OM of: US-09-303-518D-649 to: Issued_Patents_AA:* out_format : pfs

Date: Jun 30, 2002 6:57 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_n2p.model -DEV=x1h
-Q=/cgn2_1/USPTO_spool/US09303518/runat_28062002_142713_4302/app_query.fasta_1.23501
-DB=Issued_Patents_AA -QFMT=fastan -SUFFIX=rai -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=100 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=45 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09303518_@CGN1_1_279 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-303-518D-649
Query length: 4374
Database: Issued_Patents_AA:*
Database sequences: 231628
Database length: 24425594
Search time (sec): 185.890000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	! Documentation ..
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-296-791-3 +			1608.50	2101.61	9.1e-111	1541 !
/cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:PCT-US95-10661A-3 +			1608.50	2101.61	9.1e-111	1541
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-296-791-4 +			1579.50	2063.11	1.3e-108	1545 !
/cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:PCT-US95-10661A-4 +			1579.50	2063.11	1.3e-108	1545
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-296-791-5 +			1568.00	2046.78	9.3e-108	1702 !
/cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:PCT-US95-10661A-5 +			1568.00	2046.78	9.3e-108	1702
/cgn2_6/ptodata/1/iaa/backfiles1.pep:5268270-2 +			1532.50	2001.04	3.7e-105	1507 ! Pate
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-296-791-6 +			1484.00	1934.44	1.5e-101	1848 !
/cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:PCT-US95-10661A-6 +			1484.00	1934.44	1.5e-101	1848
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-460-269C-4 +			295.50	365.58	7.6e-14	911 !
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-728-470-10 +			295.50	359.84	9.5e-14	1529 !
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-719-641-10 +			295.50	359.84	9.5e-14	1529 !
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-460-269C-6 +			295.00	364.78	8.4e-14	922 !
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-617-697-10 +			292.50	355.36	1.6e-13	1600 !
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-460-269C-2 +			283.00	349.01	6.4e-13	910 !
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-728-470-9 +			281.00	342.08	1.1e-12	1338 !
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-719-641-9 +			281.00	342.08	1.1e-12	1338 !
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-617-697-9 +			281.00	340.11	1.1e-12	1599 !
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-038-682-4 +			280.50	340.32	1.2e-12	1477 !
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-302-832-4 +			280.50	340.32	1.2e-12	1477 !
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-530-198-4 +			280.50	340.32	1.2e-12	1477 !
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-469-880-4 +			280.50	340.32	1.2e-12	1477 !
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-728-470-4 +			280.50	340.32	1.2e-12	1477 !
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-617-697-4 +			280.50	340.32	1.2e-12	1477 !
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-719-641-4 +			280.50	340.32	1.2e-12	1477 !
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-038-682-2 +			272.50	329.28	4.8e-12	1536 !
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-302-832-2 +			272.50	329.28	4.8e-12	1536 !
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-530-198-2 +			272.50	329.28	4.8e-12	1536 !
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-469-880-2 +			272.50	329.28	4.8e-12	1536 !
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-728-470-2 +			272.50	329.28	4.8e-12	1536 !
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-617-697-2 +			272.50	329.28	4.8e-12	1536 !
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-719-641-2 +			272.50	329.28	4.8e-12	1536 !
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-169-927-2 +			241.00	286.95	1.0e-09	1612 !
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-268-347-36 +			240.50	281.83	1.3e-09	2411 !
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-268-347-48 +			234.00	275.01	3.8e-09	2048 !
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-268-347-49 +			232.50	271.67	5.1e-09	2314 !
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-413-814-78 +			231.50	281.03	4.0e-09	882 !
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-377-155-33 +			229.50	267.50	8.6e-09	2353 !
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-913-942-4 +			229.50	267.50	8.6e-09	2353 !
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-669-974-33 +			229.50	267.50	8.6e-09	2353 !
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-268-347-47 +			229.50	267.50	8.6e-09	2354
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-968-685A-10 +			223.00	260.02	2.5e-08	2123
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-409-995-4 +			218.50	255.21	5.1e-08	1912
/cgn2_6/ptodata/1/iaa/6A_COMB.pep:US-08-685-467-4 +			218.50	255.21	5.1e-08	1912
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-409-995-2 +			214.00	255.39	8.7e-08	1098
/cgn2_6/ptodata/1/iaa/6A_COMB.pep:US-08-685-467-2 +			214.00	255.39	8.7e-08	1098
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-377-155-32 +			214.00	255.39	8.7e-08	1098
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-913-942-2 +			214.00	255.39	8.7e-08	1098
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-669-974-32 +			214.00	255.39	8.7e-08	1098
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-268-347-44 +			214.00	255.39	8.7e-08	1098
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-074-658-11 +			213.00	245.22	1.4e-07	2439
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-074-658-15 +			207.00	237.29	4.0e-07	2432
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-413-814-79 +			204.00	241.02	5.0e-07	1213
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-387-942C-3 +			199.50	233.44	1.1e-06	1403
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-268-347-28 +			199.00	235.43	1.1e-06	1104
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-268-347-34 +			199.00	235.43	1.1e-06	1104
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-199-637A-351 +			193.50	238.15	1.9e-06	447
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-336-447A-9 +			193.00	229.24	2.9e-06	941
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-268-347-30 +			192.00	227.20	3.5e-06	1004
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-336-447A-13 +			188.00	223.44	6.6e-06	873
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-049-783-2 +			187.00	218.07	9.1e-06	1257
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-158-232-6 +			187.00	218.07	9.1e-06	1257
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-304-626-6 +			187.00	218.07	9.1e-06	1257
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-316-301A-6 +			187.00	218.07	9.1e-06	1257
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-611-928-6 +			187.00	218.07	9.1e-06	1257
/cgn2_6/ptodata/1/iaa/6A_COMB.pep:US-09-173-891-6 +			187.00	218.07	9.1e-06	1257
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-076-137-6 +			187.00	218.07	9.1e-06	1257
/cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:PCT-US92-03624-6 +			187.00	218.07	9.1e-06	12
/cgn2_6/ptodata/1/iaa/6A_COMB.pep:US-08-470-260-3 +			186.50	217.07	1.0e-05	1296
/cgn2_6/ptodata/1/iaa/6A_COMB.pep:US-08-471-491-3 +			186.50	217.07	1.0e-05	1296
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-466-662-3 +			186.50	217.07	1.0e-05	1296
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-07-741-940-7 +			186.00	207.70	1.5e-05	2842
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-289-548A-7 +			186.00	207.70	1.5e-05	2842
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-452-654-7 +			186.00	207.70	1.5e-05	2842
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-452-655B-2 +			186.00	207.70	1.5e-05	2843
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-452-655B-7 +			186.00	207.70	1.5e-05	2843
/cgn2_6/ptodata/1/iaa/6A_COMB.pep:US-08-450-582-2 +			186.00	207.70	1.5e-05	2843
/cgn2_6/ptodata/1/iaa/6A_COMB.pep:US-08-450-582-7 +			186.00	207.70	1.5e-05	2843
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-821-355A-7 +			186.00	207.20	1.6e-05	2973
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-09-003-687A-7 +			186.00	207.20	1.6e-05	2973
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-136-605-7 +			186.00	207.20	1.6e-05	2973
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-336-447A-5 +			185.50	219.88	1.0e-05	892
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-336-447A-1 +			184.50	219.34	1.2e-05	831
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-07-741-940-2 +			184.00	205.04	2.1e-05	2843
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-289-548A-2 +			184.00	205.04	2.1e-05	2843
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-452-654-2 +			184.00	205.04	2.1e-05	2843
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-370-235A-2 +			184.00	205.04	2.1e-05	2843
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-387-942C-5 +			183.50	217.48	1.4e-05	872
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-387-942C-4 +			183.50	216.00	1.5e-05	997
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-714-402-2 +			183.00	214.12	1.7e-05	1112
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-616-844-40 +			182.50	210.28	2.1e-05	1481
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-599-654-40 +			182.50	210.28	2.1e-05	1481
/cgn2_6/ptodata/1/iaa/6A_COMB.pep:US-08-944-868A-40 +			182.50	210.28	2.1e-05	1481
/cgn2_6/ptodata/1/iaa/6A_COMB.pep:US-08-944-423A-40 +			182.50	210.28	2.1e-05	1481
/cgn2_6/ptodata/1/iaa/6A_COMB.pep:US-08-944-496-40 +			182.50	210.28	2.1e-05	1481
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-327-536-2 +			182.00	212.32	2.1e-05	1161
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-268-347-32 +			179.00	209.00	3.4e-05	1094
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-336-447A-15 +			178.50	210.63	3.3e-05	889

seq_name: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-296-791-3

seq_documentation_block:
; Sequence 3, Application US/08296791
; Patent No. 6245337
; GENERAL INFORMATION:
; APPLICANT: St. Geme III, Joseph W.
; APPLICANT: Falkow, Stanley
; TITLE OF INVENTION: Haemophilus Adherence and Penetration
; TITLE OF INVENTION: Protein
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, Suite 3400

```
;       CITY:  San Francisco
;       STATE:  California
;       COUNTRY:  United States
;       ZIP:  94111-4187
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/296,791
;       FILING DATE:  25-AUG-1994
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Trecartin, Richard F.
;       REGISTRATION NUMBER:  31,801
;       REFERENCE/DOCKET NUMBER:  A-59941/RFT/RMS
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (415) 781-1989
;       TELEFAX:  (415) 398-3249
;       TELEX:  910 277299
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1541 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  unknown
US-08-296-791-3


alignment_scores:
            Quality: 1608.50              Length:   1717
              Ratio:  1.733                 Gaps:     51
 Percent Similarity:  54.048   Percent Identity: 27.257

alignment_block:
 US-09-303-518D-649 x US-08-296-791-3   ..

 Align seg 1/1  to: US-08-296-791-3  from: 1  to: 1541

      64 CGCTTCTCGCCTGCTTACTTAGCCATATGCCTGTCGTTCGGCATTCTTCC 113
         :::|||:::     :::::::|||:::   ::::::::::::::   ||
       5 LysPheLysLeuAsnPheIleAlaLeuThrValAlaTyrAlaLeuThrPr  21

     114 CCAAGCCTGGGCGGGACACACTTATTTCGGCATCAACTACCAATACTATC 163
         |   :::   |||:::   :::         ::::::||||||   :::|
      21 oTyrThrGluAlaAlaLeuValArgAspAspValAspTyrGlnIlePheA  38

     164 GCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATT 213
         ||||||||||||||||||||||||||||||:::|||||||||   ::::::
      38 rgAspPheAlaGluAsnLysGlyLysPheSerValGlyAlaThrAsnVal  54

     214 GAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAATCAATGACAAAAGC 263
           |||   :::|||::::::::::   :::|||   ::::::   :::::
      55 LeuValLysAspLysAsnAsnLysAspLeuGlyThrAlaLeuProAsnGl  71

     264 C...CCGATGATTGATTTTTCTGTGGTGTCG...CGTAACGGCGTGGCGG 307
         :   |||||||||||||||||||||||:::     :::   :::|||:
      71 yIleProMetIleAspPheSerValValAspValAspLysArgIleAlaT  88

     308 CATTGGTGGGCGATCAATATATTGTGAGCGTGGCACAT............ 345
         ::|||::::::   ||||||:::|||:::|||   |||
      88 hrLeuIleAsnProGlnTyrValValGlyValLysHisValSerAsnGly 104

     346 ..........................AACGGCGGCTATAACAACGTTGA 368
                                    ||||||:::   ||||||   ::
     105 ValSerGluLeuHisPheGlyAsnLeuAsnGlyAsnMetAsnAsnGlyAs 121

     369 TTTTGGTGCGGAAGGAAGAAATCCCGATCAACATCGTTTTACTTATAAAA 418
         :     |||:::   |||:::   ::::::::::::     |||
     121 nAlaLysAlaHis...ArgAspValSerSerGluGluAsnArgTyrPheS 137

     419 TTGTGAAACGGAATAATTATAAAGCAGGGACTAAAGGC............ 456
           |||:::::::|||:::|||   :::      :::|||

     137 erValGluLysAsnGluTyrProThrLysLeuAsnGlyLysThrValThr 153

     457 ..............CATCCTTATGGCGGCGATTATCATATGCCGCGTTT 491
                        :::             ||||||:::||||||||||||
     154 ThrGluAspGlnThrGlnLysArgArgGluAspTyrTyrMetProArgLe 170

     492 GCATAAATTTGTCACAGATGCAGAACCTGTTGAA...ATGACCAGTTATA 538
         |   ||||||||||||:::::   |||:::|||      |||:::
     170 uAspLysPheValThrGluValAlaProIleGluAlaSerThrAlaSerS 187

     539 TGGATGGGCGGAAATATATCGATCAAAATAATTACCCTGACCGTGTTCGT 588
           |||:::      |||   |||||||||:::||||||      ||||||
     187 erAspAlaGlyThrTyrAsnAspGlnAsnLysTyrProAlaPheValArg 203

     589 ATTGGGGCAGGCAGGCAATATTGGCGATCTGATGAAGATGAG........ 630
         :::|||:::|||   |||:::      :::      |||:::
     204 LeuGlySerGlySerGlnPheIleTyrLysLysGlyAspAsnTyrSerLe 220

     631 ....CCCAATAACCGCGAAAGTTCATATCAT............ATTGCAA 664
                ||||||:::|||   :::   :::             :::::::
     220 uIleLeuAsnAsnHisGluValGlyGlyAsnAsnLeuLysLeuValGlyA 237

     665 GTGCGTATTCTTGGCTCGTTGGTGGCAATACCTTTGCACAAAATGGATCA 714
         ::||||||:::::   ::::::|||:::   :::      |||   :::
     237 spAlaTyrThrTyrGlyIleAlaGlyThrProTyrLysValAsnHisGlu 253

     715 GGTGGTGGCACAGTCAACTTAGGTAGTGAAAAAATTAAACATAGC...CC 761
         :::::::|||   ::::::::::|||:::::::|||   :::||||||   ||
     254 AsnAsnGlyLeuIleGlyPheGlyAsnSerLysGluGluHisSerAspPr 270

     762 ATATGGTTTTTTACCAACAGGA..............GGCTCATTTGGCG 796
         |   |||:::|||                        :::   :::|||||
     270 oLysGlyIleLeuSerGlnAspProLeuThrAsnTyrAlaValLeuGlyA 287

     797 ACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGGTTA 846
         ||||||||||||||:::|||:::||||||   :::|||   |||||||||
     287 spSerGlySerProLeuPheValTyrAspArgGluLysGlyLysTrpLeu 303

     847 ATTAATGGGGTATTGCAAACGGGCAACCCCTATATAGGAAAAAGCAATGG 896
         :::   |||      :::             |||      ||||||
     304 PheLeuGlySerTyrAspPheTrpAlaGlyTyrAsnLysLysSer..... 318

     897 CTTCCAGCTGGTTCGTAAAGATTGG...TTCTATGATGAAATCTTTGCTG 943
                        :::::::|||   :::|||   :::   |||:::
     319 .............TrpGlnGluTrpAsnIleTyrLysSerGlnPheThrL 331

     944 GAGATACCCATTCAGTATTCTACGAACCACGTCAAAATGGGAAATACTCT 993
           |||:::
     331 ysAspVal..................................... 333

     994 TTTAACGACGATAATAATGGCACAGGAAAAATCAATGCCAAACATGAACA 1043
         :::|||   |||:::   |||:::      |||::::::|||   :::::
     334 LeuAsnLysAspSerAlaGlySer...LeuIleGlySerLysThrAspTy 349

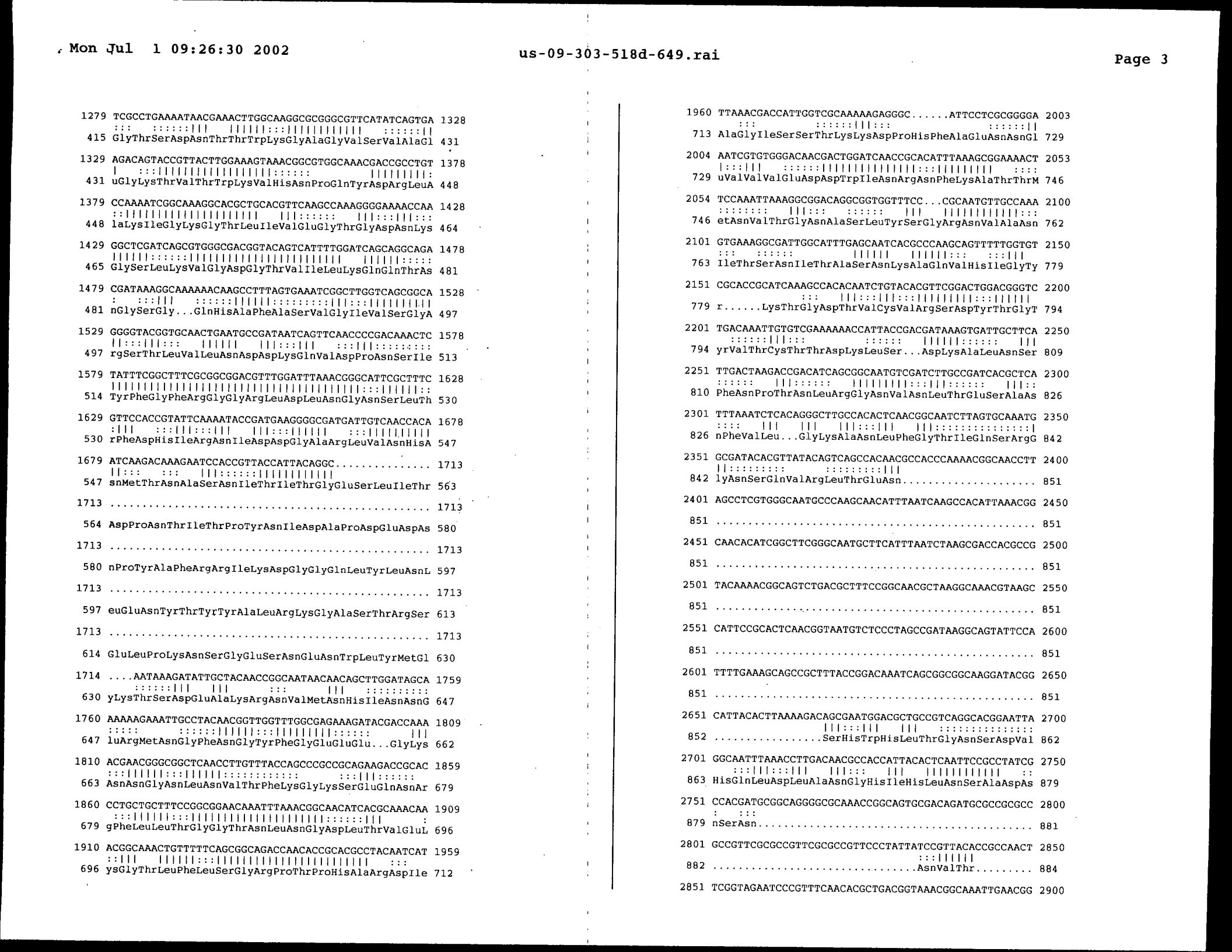
    1044 CAATTCTCTGCCTAATAGATTAAAAACACGAACCGTT...........C 1081
         ::::         |||      ||||||   |||:::          : .
     349 rSerTrpSerSerAsnGly...LysThrSerThrIleThrGlyGlyGluL 365

    1082 AATTGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTGTTTATCAT 1131
         ::   :::||||||:::|||::::::         :::|||
     365 ysSerLeuAsnValAspLeuAlaAspGlyLysAspLysPro......... 378

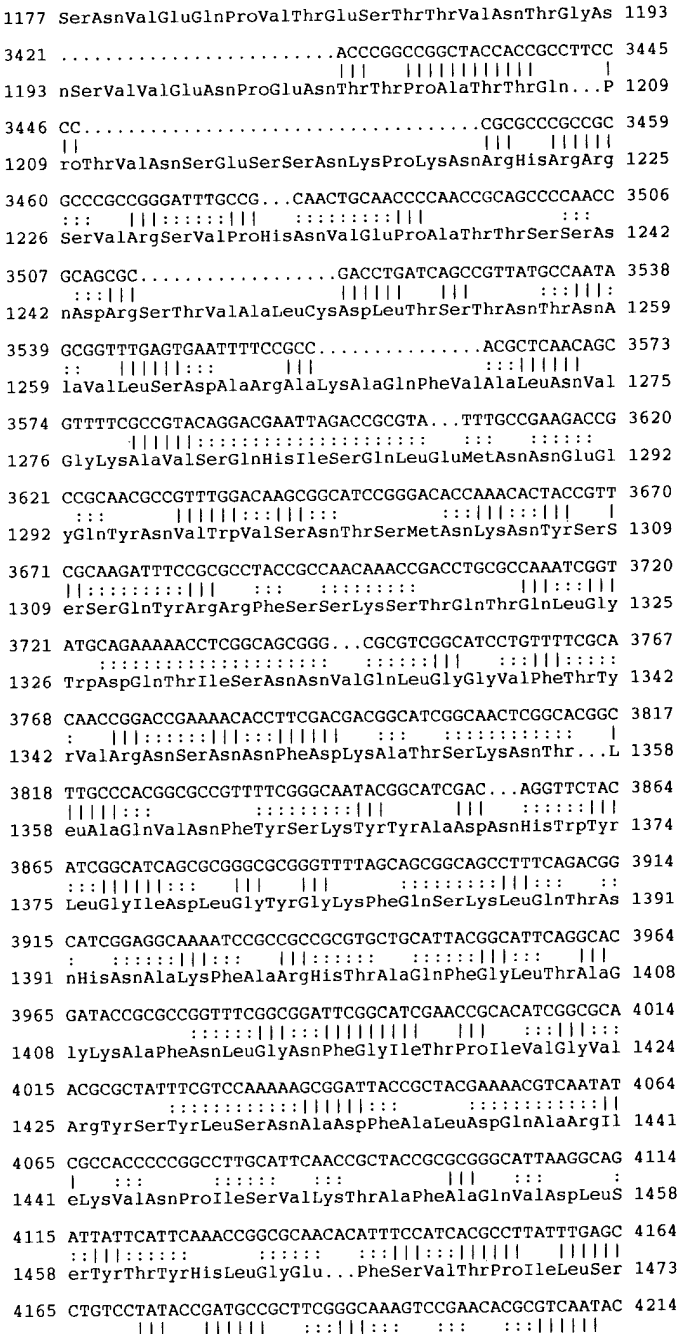
    1132 GCTGCAGGTGGTGTCAACAGTTATCGACCCAGACTGAATAATGGAGAAAA 1181
                                                |||:::|||:::::
     379 ..................................AsnHisGlyLysSe 383

    1182 TATTTCCTTTATTGACGAAGGAAAAGGCGAATTGATACTTACCAGCAACA 1231
         :::::::|||      ||||||:::|||   |||   |||::::::||||
     383 rValThrPhe......GluGlySerGlyThrLeuThrLeuAsnAsnAsnI 398

    1232 TCAATCAAGGTGCTGGAGGATTATATTTCCAAGGAGATTTTACGGTC... 1278
         ||:::|||||||||||||||||:::||||:::|||||||:::   |||
     398 leAspGlnGlyAlaGlyGlyLeuPhePheGluGlyAspTyrGluValLys 414
```


1279 TCGCCTGAAAATAACGAAACTTGGCAAGGCGCGGGCGTTCATATCAGTGA 1328
415 GlyThrSerAspAsnThrThrTrpLysGlyAlaGlyValSerValAlaGl 431
1329 AGACAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGACCGCCTGT 1378
431 uGlyLysThrValThrTrpLysValHisAsnProGlnTyrAspArgLeuA 448
1379 CCAAAATCGGCAAAGGCACGCTGCACGTTCAAGCCAAAGGGGAAAACCAA 1428
448 laLysIleGlyLysGlyThrLeuIleValGluGlyThrGlyAspAsnLys 464
1429 GGCTCGATCAGCGTGGGCGACGGTACAGTCATTTTGGATCAGCAGGCAGA 1478
465 GlySerLeuLysValGlyAspGlyThrValIleLeuLysGlnGlnThrAs 481
1479 CGATAAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCTTGGTCAGCGGCA 1528
481 nGlySerGly...GlnHisAlaPheAlaSerValGlyIleValSerGlyA 497
1529 GGGGTACGGTGCAACTGAATGCCGATAATCAGTTCAACCCCGACAAACTC 1578
497 rgSerThrLeuValLeuAsnAspAspLysGlnValAspProAsnSerIle 513
1579 TATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAACGGGCATTCGCTTTC 1628
514 TyrPheGlyPheArgGlyGlyArgLeuAspLeuAsnGlyAsnSerLeuTh 530
1629 GTTCCACCGTATTCAAAATACCGATGAAGGGGCGATGATTGTCAACCACA 1678
530 rPheAspHisIleArgAsnIleAspAspGlyAlaArgLeuValAsnHisA 547
1679 ATCAAGACAAAGAATCCACCGTTACCATTACAGGC..... 1713
547 snMetThrAsnAlaSerAsnIleThrIleThrGlyGluSerLeuIleThr 563
1713 1713
564 AspProAsnThrIleThrProTyrAsnIleAspAlaProAspGluAspAs 580
1713 1713
580 nProTyrAlaPheArgArgIleLysAspGlyGlyGlnLeuTyrLeuAsnL 597
1713 1713
597 euGluAsnTyrThrTyrTyrAlaLeuArgLysGlyAlaSerThrArgSer 613
1713 1713
614 GluLeuProLysAsnSerGlyGluSerAsnGluAsnTrpLeuTyrMetGl 630
1714 AATAAAGATATTGCTACAACCGGCAATAACAACAGCTTGGATAGCA 1759
630 yLysThrSerAspGluAlaLysArgAsnValMetAsnHisIleAsnAsnG 647
1760 AAAAAGAAATTGCCTACAACGGTTGGTTTGGCGAGAAAGATACGACCAAA 1809
647 luArgMetAsnGlyPheAsnGlyTyrPheGlyGluGluGlu...GlyLys 662
1810 ACGAACGGGCGGCTCAACCTTGTTTACCAGCCCGCCGCAGAAGACCGCAC 1859
663 AsnAsnGlyAsnLeuAsnValThrPheLysGlyLysSerGluGlnAsnAr 679
1860 CCTGCTGCTTTCCGGCGGAACAAATTTAAACGGCAACATCACGCAAACAA 1909
679 gPheLeuLeuThrGlyGlyThrAsnLeuAsnGlyAspLeuThrValGluL 696
1910 ACGGCAAACTGTTTTTCAGCGGCAGACCAACACCGCACGCCTACAATCAT 1959
696 ysGlyThrLeuPheLeuSerGlyArgProThrProHisAlaArgAspIle 712
1960 TTAAACGACCATTGGTCGCAAAAAGAGGGC......ATTCCTCGCGGGGA 2003
713 AlaGlyIleSerSerThrLysLysAspProHisPheAlaGluAsnAsnGl 729
2004 AATCGTGTGGGACAACGACTGGATCAACCGCACATTTAAAGCGGAAAACT 2053
729 uValValValGluAspAspTrpIleAsnArgAsnPheLysAlaThrThrM 746
2054 TCCAAATTAAAGGCGGACAGGCGGTGGTTTCC...CGCAATGTTGCCAAA 2100
746 etAsnValThrGlyAsnAlaSerLeuTyrSerGlyArgAsnValAlaAsn 762
2101 GTGAAAGGCGATTGGCATTTGAGCAATCACGCCCAAGCAGTTTTTGGTGT 2150
763 IleThrSerAsnIleThrAlaSerAsnLysAlaGlnValHisIleGlyTy 779
2151 CGCACCGCATCAAAGCCACACAATCTGTACACGTTCGGACTGGACGGGTC 2200
779 r......LysThrGlyAspThrValCysValArgSerAspTyrThrGlyT 794
2201 TGACAAATTGTGTCGAAAAAACCATTACCGACGATAAAGTGATTGCTTCA 2250
794 yrValThrCysThrThrAspLysLeuSer...AspLysAlaLeuAsnSer 809
2251 TTGACTAAGACCGACATCAGCGGCAATGTCGATCTTGCCGATCACGCTCA 2300
810 PheAsnProThrAsnLeuArgGlyAsnValAsnLeuThrGluSerAlaAs 826
2301 TTTAAATCTCACAGGGCTTGCCACACTCAACGGCAATCTTAGTGCAAATG 2350
826 nPheValLeu...GlyLysAlaAsnLeuPheGlyThrIleGlnSerArgG 842
2351 GCGATACACGTTATACAGTCAGCCACAACGCCACCCAAAACGGCAACCTT 2400
842 lyAsnSerGlnValArgLeuThrGluAsn..... 851
2401 AGCCTCGTGGGCAATGCCCAAGCAACATTTAATCAAGCCACATTAAACGG 2450
851 851
2451 CAACACATCGGCTTCGGGCAATGCTTCATTTAATCTAAGCGACCACGCCG 2500
851 851
2501 TACAAAACGGCAGTCTGACGCTTTCCGGCAACGCTAAGGCAAACGTAAGC 2550
851 851
2551 CATTCCGCACTCAACGGTAATGTCTCCCTAGCCGATAAGGCAGTATTCCA 2600
851 851
2601 TTTTGAAAGCAGCCGCTTTACCGGACAAATCAGCGGCGGCAAGGATACGG 2650
851 851
2651 CATTACACTTAAAAGACAGCGAATGGACGCTGCCGTCAGGCACGGAATTA 2700
852 SerHisTrpHisLeuThrGlyAsnSerAspVal 862
2701 GGCAATTTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCG 2750
863 HisGlnLeuAspLeuAlaAsnGlyHisIleHisLeuAsnSerAlaAspAs 879
2751 CCACGATGCGGCAGGGGCGCAAACCGGCAGTGCGACAGATGCGCCGCGCC 2800
879 nSerAsn..... 881
2801 GCCGTTCGCGCCGTTCGCGCCGTTCCCTATTATCCGTTACACCGCCAACT 2850
882 AsnValThr..... 884
2851 TCGGTAGAATCCCGTTTCAACACGCTGACGGTAAACGGCAAATTGAACGG 2900


885 LysTyrAsnThrLeuThrValAsn...SerLeuSerGl 896
2901 TCAGGGAACATTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGACA 2950
896 yAsnGlySerPheTyrTyrLeuThrAspLeuSerAsnLysGlnGlyAspL 913
2951 AATTGAAGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCGGTCAAC 3000
913 ysValValValThrLysSerAlaThrGlyAsnPheThrLeuGlnValAla 929
3001 AATACCGGCAACGAACCTGCAAGCCTCGAACAATTGACGGTAGTGGAAGG 3050
930 AspLysThrGlyGluPro...AsnHisAsnGluLeuThrLeuPheAspAl 945
3051 AAAAGACAACAAACCGCTGTCCGAAAACCTTAATTTCACCCTGCAAAACG 3100
945 a......SerLysAlaGlnArgAspHisLeuAsnValSerLeuValGlyA 960
3101 AACACGTCGATGCCGGCGCGTGGCGTTACCAACTCATCCGCAAAGACGGC 3150
960 snThrValAspLeuGlyAlaTrpLysTyrLysLeuArgAsnValAsnGly 976
3151 GAGTTCCGCCTGCATAATCCGGTCAAAGAACAA..... 3183
977 ArgTyrAspLeuTyrAsnProGluValGluLysArgAsnGlnThrValAs 993
3183 3183
993 pThrThrAsnIleThrThrProAsnAsnIleGlnAlaAspValProSerV 1010
3184 GAGCTTTCCGACAAACTCGGCAAG..... 3207
1010 alProSerAsnAsnGluGluIleAlaArgValAspGluAlaProValPro 1026
3208 GCAGAAGC 3215
1027 ProProAlaProAlaThrProSerGluThrThrGluThrValAlaGluAs 1043
3216 CAAAAAACAGGCGGAAAAA.....GACAACGCGCAAAGCCTTGACG 3256
1043 nSerLysGlnGluSerLysThrValGluLysAsnGluGlnAspAlaThrG 1060
3257 CGCTGATTGCGGCCGGGCGCGATGCCGTCGAA..... 3288
1060 luThrThrAlaGlnAsnArgGluValAlaLysGluAlaLysSerAsnVal 1076
3289 AAGACAGAAAGCGTTGCCGAACCGGCCCGGCAGGCAGGCGGGGAAAATGT 3338
1077 LysAlaAsnThrGlnThrAsnGluValAlaGlnSerGlySerGluThrLy 1093
3339 CGGCATTATGCAGGCGGAG.....GAAG 3361
1093 sGluThrGlnThrThrGluThrLysGluThrAlaThrValGluLysGluG 1110
3362 AGAAAAAACGGGTGCAGGCGGATAAAGACACCGCCTTGGCGAAA..... 3405
1110 luLysAlaLysValGluThrGluLysThrGlnGluValProLysValThr 1126
3406 CAGCGCGAAGC 3416
1127 SerGlnValSerProLysGlnGluGlnSerGluThrValGlnProGlnAl 1143
3417 GGAA..... 3420
1143 aGluProAlaArgGluAsnAspProThrValAsnIleLysGluProGlnS 1160
3420 3420
1160 erGlnThrAsnThrThrAlaAspThrGluGlnProAlaLysGluThrSer 1176
3420 3420


1177 SerAsnValGluGlnProValThrGluSerThrThrValAsnThrGlyAs 1193
3421 ACCCGGCCGGCTACCACCGCCTTCC 3445
1193 nSerValValGluAsnProGluAsnThrThrProAlaThrThrGln...P 1209
3446 CC.....CGCGCCCGCCGC 3459
1209 roThrValAsnSerGluSerSerAsnLysProLysAsnArgHisArgArg 1225
3460 GCCCGCCGGGATTTGCCG...CAACTGCAACCCCAACCGCAGCCCCAACC 3506
1226 SerValArgSerValProHisAsnValGluProAlaThrThrSerSerAs 1242
3507 GCAGCGC.....GACCTGATCAGCCGTTATGCCAATA 3538
1242 nAspArgSerThrValAlaLeuCysAspLeuThrSerThrAsnThrAsnA 1259
3539 GCGGTTTGAGTGAATTTTCCGCC.....ACGCTCAACAGC 3573
1259 laValLeuSerAspAlaArgAlaLysAlaGlnPheValAlaLeuAsnVal 1275
3574 GTTTTCGCCGTACAGGACGAATTAGACCGCGTA...TTTGCCGAAGACCG 3620
1276 GlyLysAlaValSerGlnHisIleSerGlnLeuGluMetAsnAsnGluGl 1292
3621 CCGCAACGCCGTTTGGACAAGCGGCATCCGGGACACCAAACACTACCGTT 3670
1292 yGlnTyrAsnValTrpValSerAsnThrSerMetAsnLysAsnTyrSerS 1309
3671 CGCAAGATTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCAAATCGGT 3720
1309 erSerGlnTyrArgArgPheSerSerLysSerThrGlnThrGlnLeuGly 1325
3721 ATGCAGAAAAACCTCGGCAGCGGG...CGCGTCGGCATCCTGTTTTCGCA 3767
1326 TrpAspGlnThrIleSerAsnAsnValGlnLeuGlyGlyValPheThrTy 1342
3768 CAACCGGACCGAAAACACCTTCGACGACGGCATCGGCAACTCGGCACGGC 3817
1342 rValArgAsnSerAsnAsnPheAspLysAlaThrSerLysAsnThr...L 1358
3818 TTGCCCACGGCGCCGTTTTCGGGCAATACGGCATCGAC...AGGTTCTAC 3864
1358 euAlaGlnValAsnPheTyrSerLysTyrTyrAlaAspAsnHisTrpTyr 1374
3865 ATCGGCATCAGCGCGGGCGCGGGTTTTAGCAGCGGCAGCCTTTCAGACGG 3914
1375 LeuGlyIleAspLeuGlyTyrGlyLysPheGlnSerLysLeuGlnThrAs 1391
3915 CATCGGAGGCAAAATCCGCCGCCGCGTGCTGCATTACGGCATTCAGGCAC 3964
1391 nHisAsnAlaLysPheAlaArgHisThrAlaGlnPheGlyLeuThrAlaG 1408
3965 GATACCGCGCCGGTTTCGGCGGATTCGGCATCGAACCGCACATCGGCGCA 4014
1408 lyLysAlaPheAsnLeuGlyAsnPheGlyIleThrProIleValGlyVal 1424
4015 ACGCGCTATTTCGTCCAAAAAGCGGATTACCGCTACGAAAACGTCAATAT 4064
1425 ArgTyrSerTyrLeuSerAsnAlaAspPheAlaLeuAspGlnAlaArgIl 1441
4065 CGCCACCCCCGGCCTTGCATTCAACCGCTACCGCGCGGGCATTAAGGCAG 4114
1441 eLysValAsnProIleSerValLysThrAlaPheAlaGlnValAspLeuS 1458
4115 ATTATTCATTCAAACCGGCGCAACACATTTCCATCACGCCTTATTTGAGC 4164
1458 erTyrThrTyrHisLeuGlyGlu...PheSerValThrProIleLeuSer 1473
4165 CTGTCCTATACCGATGCCGCTTCGGGCAAAGTCCGAACACGCGTCAATAC 4214
1474 AlaArgTyr...AspAlaAsnGlnGlySerGlyLysIleAsnValAsnGl 1489

```
     4215 CGCCGTATTGGCTCAGGATTTCGGCAAAACCCGCAGTGCGGAATGGGGCG 4264
     1489 yTyrAspPheAlaTyrAsnValGluAsnGlnGlnGlnTyrAsnAlaGlyL 1506

     4265 TAAACGCCGAAATCAAAGGTTTCACGCTGTCCCTCCACGCTGCCGCCGCC 4314
     1506 euLysLeuLysTyrHisAsnValLysLeuSerLeuIleGlyGlyLeuThr 1522

     4315 AAAGGCCCGCAACTGGAAGCGCAACACAGCGCGGGCATCAAATTAGGCTA 4364
     1523 LysAlaLysGlnAlaGluLysGlnLysThrAlaGluLeuLysLeuSerPh 1539

     4365 C 4365
          :
     1539 e 1539

seq_name: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:PCT-US95-10661A-3

seq_documentation_block:
; Sequence 3, Application PC/TUS9510661A
;  GENERAL INFORMATION:
;     APPLICANT:  Washington University, et al.
;     TITLE OF INVENTION:  Haemophilus Adherence and Penetration Protein
;     NUMBER OF SEQUENCES:  9
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Flehr, Hohbach, Test, Albritton & Herbert
;       STREET:  4 Embarcadero Center, Suite 3400
;       CITY:  San Francisco
;       STATE:  California
;       COUNTRY:  United States
;       ZIP:  94111-4187
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US95/10661A
;       FILING DATE:  16-AUG-1995
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/296,791
;       FILING DATE:  25-AUG-1994
;       CLASSIFICATION:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Trecartin, Richard F.
;       REGISTRATION NUMBER:  31,801
;       REFERENCE/DOCKET NUMBER:  FP-59941/RFT
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (415) 781-1989
;       TELEFAX:  (415) 398-3249
;       TELEX:  910 277299
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1541 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  unknown
PCT-US95-10661A-3


alignment_scores:
            Quality: 1608.50              Length:   1717
              Ratio:  1.733                Gaps:     51
 Percent Similarity:  54.048   Percent Identity: 27.257

alignment_block:
 US-09-303-518D-649 x PCT-US95-10661A-3   ..

 Align seg 1/1  to: PCT-US95-10661A-3  from: 1  to: 1541

      64 CGCTTCTCGCCTGCTTACTTAGCCATATGCCTGTCGTTCGGCATTCTTCC 113

       5 LysPheLysLeuAsnPheIleAlaLeuThrValAlaTyrAlaLeuThrPr 21

     114 CCAAGCCTGGGCGGGACACACTTATTTCGGCATCAACTACCAATACTATC 163
      21 oTyrThrGluAlaAlaLeuValArgAspAspValAspTyrGlnIlePheA 38

     164 GCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATT 213
      38 rgAspPheAlaGluAsnLysGlyLysPheSerValGlyAlaThrAsnVal 54

     214 GAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAATCAATGACAAAAGC 263
      55 LeuValLysAspLysAsnAsnLysAspLeuGlyThrAlaLeuProAsnGl 71

     264 C...CCGATGATTGATTTTTCTGTGGTGTCG...CGTAACGGCGTGGCGG 307
      71 yIleProMetIleAspPheSerValValAspValAspLysArgIleAlaT 88

     308 CATTGGTGGGCGATCAATATATTGTGAGCGTGGCACAT..... 345
      88 hrLeuIleAsnProGlnTyrValValGlyValLysHisValSerAsnGly 104

     346 .....AACGGCGGCTATAACAACGTTGA 368
     105 ValSerGluLeuHisPheGlyAsnLeuAsnGlyAsnMetAsnAsnGlyAs 121

     369 TTTTGGTGCGGAAGGAAGAAATCCCGATCAACATCGTTTTACTTATAAAA 418
     121 nAlaLysAlaHis...ArgAspValSerSerGluGluAsnArgTyrPheS 137

     419 TTGTGAAACGGAATAATTATAAAGCAGGGACTAAAGGC..... 456
     137 erValGluLysAsnGluTyrProThrLysLeuAsnGlyLysThrValThr 153

     457 .....CATCCTTATGGCGGCGATTATCATATGCCGCGTTT 491
     154 ThrGluAspGlnThrGlnLysArgArgGluAspTyrTyrMetProArgLe 170

     492 GCATAAATTTGTCACAGATGCAGAACCTGTTGAA...ATGACCAGTTATA 538
     170 uAspLysPheValThrGluValAlaProIleGluAlaSerThrAlaSerS 187

     539 TGGATGGGCGGAAATATATCGATCAAAATAATTACCCTGACCGTGTTCGT 588
     187 erAspAlaGlyThrTyrAsnAspGlnAsnLysTyrProAlaPheValArg 203

     589 ATTGGGGCAGGCAGGCAATATTGGCGATCTGATGAAGATGAG..... 630
     204 LeuGlySerGlySerGlnPheIleTyrLysLysGlyAspAsnTyrSerLe 220

     631 ....CCCAATAACCGCGAAAGTTCATATCAT.....ATTGCAA 664
     220 uIleLeuAsnAsnHisGluValGlyGlyAsnAsnLeuLysLeuValGlyA 237

     665 GTGCGTATTCTTGGCTCGTTGGTGGCAATACCTTTGCACAAAATGGATCA 714
     237 spAlaTyrThrTyrGlyIleAlaGlyThrProTyrLysValAsnHisGlu 253

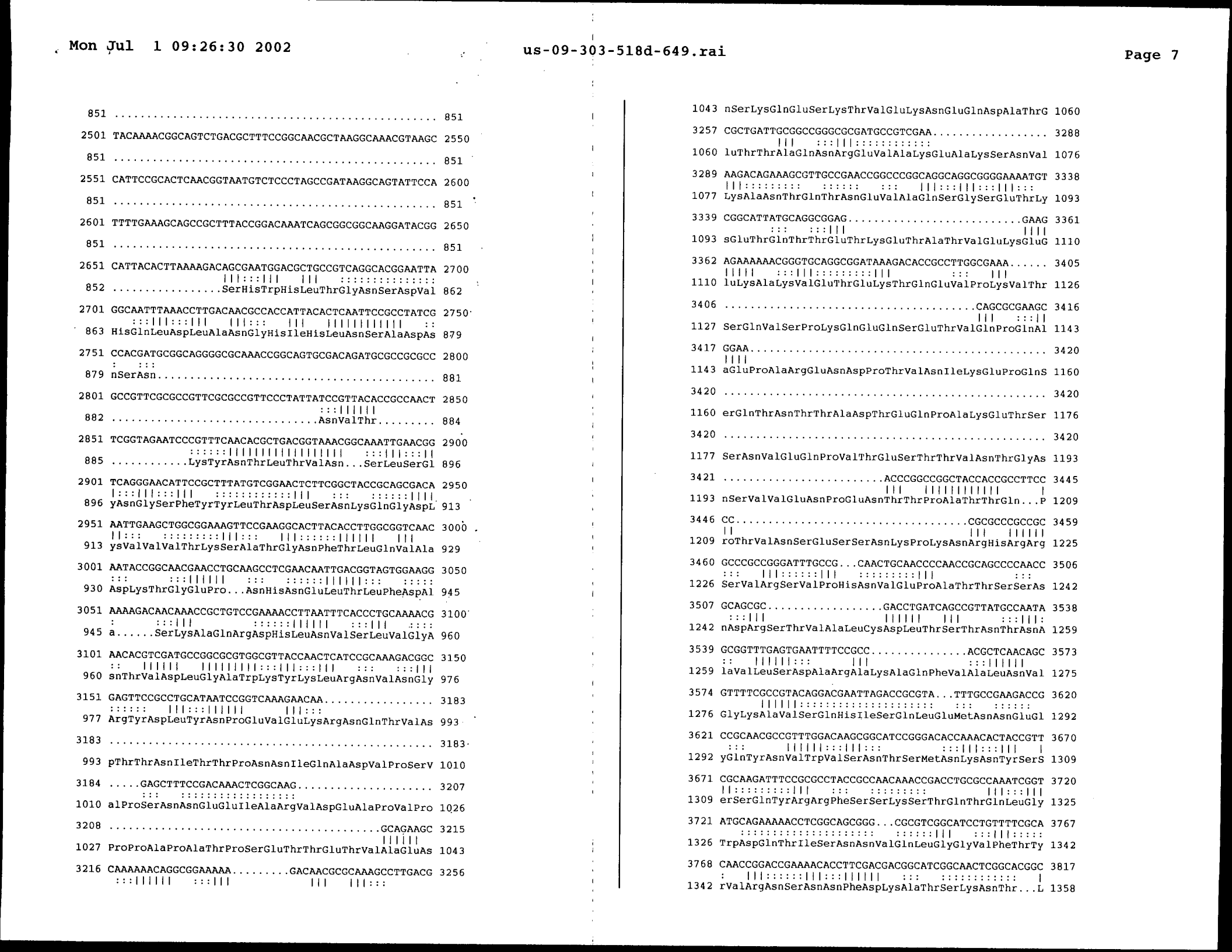
     715 GGTGGTGGCACAGTCAACTTAGGTAGTGAAAAAATTAAACATAGC...CC 761
     254 AsnAsnGlyLeuIleGlyPheGlyAsnSerLysGluGluHisSerAspPr 270

     762 ATATGGTTTTTTACCAACAGGA.....GGCTCATTTGGCG 796
     270 oLysGlyIleLeuSerGlnAspProLeuThrAsnTyrAlaValLeuGlyA 287

     797 ACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGGTTA 846
     287 spSerGlySerProLeuPheValTyrAspArgGluLysGlyLysTrpLeu 303

     847 ATTAATGGGGTATTGCAAACGGGCAACCCCTATATAGGAAAAAGCAATGG 896
     304 PheLeuGlySerTyrAspPheTrpAlaGlyTyrAsnLysLysSer..... 318
```





```
   3818 TTGCCCACGGCGCCGTTTTCGGGCAATACGGCATCGAC...AGGTTCTAC 3864
        ||||||::         :::::::::|||      |||   :::::::|||
   1358 euAlaGlnValAsnPheTyrSerLysTyrTyrAlaAspAsnHisTrpTyr 1374

   3865 ATCGGCATCAGCGCGGGCGCGGGTTTTAGCAGCGGCAGCCTTTCAGACGG 3914
        :::||||||:::   |||   |||      ::::::::::|||:::   ::
   1375 LeuGlyIleAspLeuGlyTyrGlyLysPheGlnSerLysLeuGlnThrAs 1391

   3915 CATCGGAGGCAAAATCCGCCGCCGCGTGCTGCATTACGGCATTCAGGCAC 3964
        :   :::::|||:::   |||::::::   ::::::|||:::   |||
   1391 nHisAsnAlaLysPheAlaArgHisThrAlaGlnPheGlyLeuThrAlaG 1408

   3965 GATACCGCGCCGGTTTCGGCGGATTCGGCATCGAACCGCACATCGGCGCA 4014
                ::::::|||:::|||||||||   |||   :::|||:::
   1408 lyLysAlaPheAsnLeuGlyAsnPheGlyIleThrProIleValGlyVal 1424

   4015 ACGCGCTATTTCGTCCAAAAAGCGGATTACCGCTACGAAAACGTCAATAT 4064
              :::::::::::::||||||:::      :::::::::::||
   1425 ArgTyrSerTyrLeuSerAsnAlaAspPheAlaLeuAspGlnAlaArgIl 1441

   4065 CGCCACCCCCGGCCTTGCATTCAACCGCTACCGCGCGGGCATTAAGGCAG 4114
        |   :::      ::::::   :::         |||   :::      :
   1441 eLysValAsnProIleSerValLysThrAlaPheAlaGlnValAspLeuS 1458

   4115 ATTATTCATTCAAACCGGCGCAACACATTTCCATCACGCCTTATTTGAGC 4164
        ::|||:::::      ::::::   :::|||:::||||||   ||||||
   1458 erTyrThrTyrHisLeuGlyGlu...PheSerValThrProIleLeuSer 1473

   4165 CTGTCCTATACCGATGCCGCTTCGGGCAAAGTCCGAACACGCGTCAATAC 4214
              |||   ||||||   :::|||:::   :::   :::||||||
   1474 AlaArgTyr...AspAlaAsnGlnGlySerGlyLysIleAsnValAsnGl 1489

   4215 CGCCGTATTGGCTCAGGATTTCGGCAAAACCCGCAGTGCGGAATGGGGCG 4264
              :::|||   :::      :::   ::::::   :::   |||:
   1489 yTyrAspPheAlaTyrAsnValGluAsnGlnGlnGlnTyrAsnAlaGlyL 1506

   4265 TAAACGCCGAAATCAAAGGTTTCACGCTGTCCCTCCACGCTGCCGCCGCC 4314
        :::::   :::      :::      |||||||||   ::::::   :::
   1506 euLysLeuLysTyrHisAsnValLysLeuSerLeuIleGlyGlyLeuThr 1522

   4315 AAAGGCCCGCAACTGGAAGCGCAACACAGCGCGGGCATCAAATTAGGCTA 4364
        |||:::   |||   |||   |||   :::|||   :::||||||:::::
   1523 LysAlaLysGlnAlaGluLysGlnLysThrAlaGluLeuLysLeuSerPh 1539

   4365 C 4365
        :
   1539 e 1539

seq_name: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-296-791-4

seq_documentation_block:
; Sequence 4, Application US/08296791
; Patent No. 6245337
;  GENERAL INFORMATION:
;    APPLICANT:  St. Geme III, Joseph W.
;    APPLICANT:  Falkow, Stanley
;    TITLE OF INVENTION:  Haemophilus Adherence and Penetration
;    TITLE OF INVENTION:  Protein
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Flehr, Hohbach, Test, Albritton & Herbert
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/296,791
;      FILING DATE:  25-AUG-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Trecartin, Richard F.
;      REGISTRATION NUMBER:  31,801
;      REFERENCE/DOCKET NUMBER:  A-59941/RFT/RMS
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1545 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
US-08-296-791-4


alignment_scores:
            Quality: 1579.50              Length:   1735
              Ratio:  1.704                 Gaps:     53
 Percent Similarity:  53.429   Percent Identity: 26.686

alignment_block:
 US-09-303-518D-649 x US-08-296-791-4   ..

 Align seg 1/1  to: US-08-296-791-4  from: 1  to: 1545

     64 CGCTTCTCGCCTGCTTACTTAGCCATATGCCTGTCGTTCGGCATTCTTCC 113
        :::|||:::      :::::|||:::   ::::::::::::::   ||
      5 LysPheLysLeuAsnPheIleAlaLeuThrValAlaTyrAlaLeuThrPr 21

    114 CCAAGCCTGGGCGGGACACACTTATTTCGGCATCAACTACCAATACTATC 163
        |   :::   |||:::   :::         ::::::||||||   :::|
     21 oTyrThrGluAlaAlaLeuValArgAspAspValAspTyrGlnIlePheA 38

    164 GCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATT 213
        ||||||||||||||||||||||||||||||:::|||||||||   ::::::
     38 rgAspPheAlaGluAsnLysGlyLysPheSerValGlyAlaThrAsnVal 54

    214 GAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAATCAATGACAAAAGC 263
        ||||||   :::|||:::::::::   :::|||:::   :::   :::::
     55 GluValArgAspLysAsnAsnArgProLeuGlyAsnValLeuProAsnGl 71

    264 C...CCGATGATTGATTTTTCTGTGGTGTCG...CGTAACGGCGTGGCGG 307
        :   |||||||||||||||||||||||:::      :::   :::|||:
     71 yIleProMetIleAspPheSerValValAspValAspLysArgIleAlaT 88

    308 CATTGGTGGGCGATCAATATATTGTGAGCGTGGCACAT..... 345
        ::||||||:::   ||||||:::|||:::|||   |||
     88 hrLeuValAsnProGlnTyrValValGlyValLysHisValSerAsnGly 104

    346 .....AACGGCGGCTATAACAACGTTGA 368
                ||||||:::   ||||||   ::
    105 ValSerGluLeuHisPheGlyAsnLeuAsnGlyAsnMetAsnAsnGlyAs 121

    369 TTTTGGTGCGGAAGGAAGAAATCCCGATCAACATCGTTTTACTTATAAAA 418
        :      |||:::   |||:::   ::::::::::::      |||
    121 nAlaLysAlaHis...ArgAspValSerSerGluGluAsnArgTyrTyrT 137

    419 TTGTGAAACGGAATAATTAT.................AAAGCAGGGACT 450
          |||:::::::|||:::|||                ||||||   |||
    137 hrValGluLysAsnGluTyrProThrLysLeuAsnGlyLysAlaValThr 153

    451 .........AAAGGCCATCCTTATGGCGGCGATTATCATATGCCGCGTTT 491
                :::::::::              ||||||:::||||||||||
    154 ThrGluAspGlnAlaGlnLysArgArgGluAspTyrTyrMetProArgLe 170

    492 GCATAAATTTGTCACAGATGCAGAACCTGTTGAA...ATGACCAGTTATA 538
        |   ||||||||||||:::::   |||:::|||      |||:::
    170 uAspLysPheValThrGluValAlaProIleGluAlaSerThrAspSerS 187
```

```
539 TGGATGGGCGGAAATATATCGATCAAAATAATTACCCTGACCGTGTTCGT 588
187 erThrAlaGlyThrTyrAsnAsnLysAspLysTyrProTyrPheValArg 203

589 ATTGGGGCAGGCAGGCAATAT..............................  609
204 LeuGlySerGlyThrGlnPheIleTyrGluAsnGlyThrArgTyrGluLe 220

610 .TGGCGATCTGATGAAGATGAGCCCAATAACCGCGAAAGTTCATATCAT.  657
220 uTrpLeuGlyLysGluGlyGlnLysSerAsp...AlaGlyGlyTyrAsnL 236

658 ........ATTGCAAGTGCGTATTCTTGGCTCGTTGGTGGCAATACCTTT 699
236 euLysLeuValGlyAsnAlaTyrThrTyrGlyIleAlaGlyThrProTyr 252

700 GCACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGT........  741
253 GluValAsnHisGluAsnAspGlyLeuIleGlyPheGlyAsnSerAsnAs 269

742 ..................GAAAAAATTAAACATAGCCCATATGGTTTTT 772
269 nGluTyrIleAsnProLysGluIleLeuSerLysLysPro.........L 283

773 TACCAACAGGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATCTAT 822
283 euThrAsnTyrAlaValLeuGlyAspSerGlySerProLeuPheValTyr 299

823 GATGCCCAAAAGCAAAAGTGGTTAATTAATGGGGTATTGCAAACGGGCAA 872
300 AspArgGluLysGlyLysTrpLeuPheLeuGlySerTyrAspTyrTrpAl 316

873 CCCCTATATAGGAAAAAGCAATGGCTTCCAGCTGGTTCGTAAAGATTGGT 922
316 aGlyTyr..........................AsnLysLysSerTrpG 324

923 TCTATGATGAAATCTTTGCTGGAGATACCCATTCAGTATTCTACGAACCA 972
324 lnGluTrpAsnIleTyrLysProGluPheAlaGluLysIleTyrGluGln 340

973 CGTCAAAATGGG..............AAATACTCTTTTAACGACGATAA 1007
341 TyrSerAlaGlySerLeuIleGlySerLysThrAspTyrSerTrpSerSe 357

1008 TAATGGCACAGGAAAAATCAATGCCAAACATGAACACAATTCTCTGCCTA 1057
357 rAsnGly.......................................... 359

1058 ATAGATTAAAAACACGAACCGTT............CAATTGTTTAATGTT 1095
360 ........LysThrSerThrIleThrGlyGlyGluLysSerLeuAsnVal 373

1096 TCTTTATCCGAGACAGCAAGAGAACCTGTTTATCATGCTGCAGGTGGTGT 1145
374 AspLeuAlaAspGlyLysAspLysPro....................... 382

1146 CAACAGTTATCGACCCAGACTGAATAATGGAGAAAATATTTCCTTTATTG 1195
383 .....................AsnHisGlyLysSerValThrPhe.... 390

1196 ACGAAGGAAAAGGCGAATTGATACTTACCAGCAACATCAATCAAGGTGCT 1245
391 ..GluGlySerGlyThrLeuThrLeuAsnAsnAsnIleAspGlnGlyAla 406

1246 GGAGGATTATATTTCCAAGGAGATTTTACGGTC...TCGCCTGAAAATAA 1292
407 GlyGlyLeuPheGluGlyAspTyrGluValLysGlyThrSerAspAs 423

1293 CGAAACTTGGCAAGGCGCGGGCGTTCATATCAGTGAAGACAGTACCGTTA 1342
423 nThrThrTrpLysGlyAlaGlyValSerValAlaGluGlyLysThrValT 440

1343 CTTGGAAAGTAAACGGCGTGGCAAACGACCGCCTGTCCAAAATCGGCAAA 1392

1393 GGCACGCTGCACGTTCAAGCCAAAGGGGAAAACCAAGGCTCGATCAGCGT 1442
457 GlyThrLeuIleValGluGlyThrGlyAspAsnLysGlySerLeuLysVa 473

1443 GGGCGACGGTACAGTCATTTTGGATCAGCAGGCAGACGATAAAGGCAAAA 1492
473 lGlyAspGlyThrValIleLeuLysGlnGlnThrAsnGlySerGly...G 489

1493 AACAAGCCTTTAGTGAAATCGGCTTGGTCAGCGGCAGGGGTACGGTGCAA 1542
489 lnHisAlaPheAlaSerValGlyIleValSerGlyArgSerThrLeuVal 505

1543 CTGAATGCCGATAATCAGTTCAACCCCGACAAACTCTATTTCGGCTTTCG 1592
506 LeuAsnAspAspLysGlnValAspProAsnSerIleTyrPheGlyPheAr 522

1593 CGGCGGACGTTTGGATTTAAACGGGCATTCGCTTTCGTTCCACCGTATTC 1642
522 gGlyGlyArgLeuAspLeuAsnGlyAsnSerLeuThrPheAspHisIleA 539

1643 AAAATACCGATGAAGGGGCGATGATTGTCAACCACAATCAAGACAAAGAA 1692
539 rgAsnIleAspGluGlyAlaArgLeuValAsnHisSerThrSerLysHis 555

1693 TCCACCGTTACCATTACAGGCAATAAAGATATTGCTACAACCGGCAATAA 1742
556 SerThrValThrIleThrGlyAsp...AsnLeuIleThrAspProAsnAs 571

1743 C.................................................. 1743
571 nValSerIleTyrTyrValLysProLeuGluAspAspAsnProTyrAlaI 588

1743 .................................................. 1743
588 leArgGlnIleLysTyrGlyTyrGlnLeuTyrPheAsnGluGluAsnArg 604

1743 .................................................. 1743
605 ThrTyrTyrAlaLeuLysLysAspAlaSerIleArgSerGluPheProGl 621

1743 .................................................. 1743
621 nAsnArgGlyGluSerAsnAsnSerTrpLeuTyrMetGlyThrGluLysA 638

1744 ......................AACAGCTTGGATAGCAAAAAAGAAATT 1770
638 laAspAlaGlnLysAsnAlaMetAsnHisIleAsnAsnGluArgMetAsn 654

1771 GCCTACAACGGTTGGTTTGGCGAGAAAGATACGACCAAAACGAACGGGCG 1820
655 GlyPheAsnGlyTyrPheGlyGluGluGlu...GlyLysAsnAsnGlyAs 670

1821 GCTCAACCTTGTTTACCAGCCCGCCGCAGAAGACCGCACCCTGCTGCTTT 1870
670 nLeuAsnValThrPheLysGlyLysSerGluGlnAsnArgPheLeuLeuT 687

1871 CCGGCGGAACAAATTTAAACGGCAACATCACGCAAACAAACGGCAAACTG 1920
687 hrGlyGlyThrAsnLeuAsnGlyAspLeuAsnValGlnGlnGlyThrLeu 703

1921 TTTTTCAGCGGCAGACCAACACCGCACGCCTACAAT.............. 1956
704 PheLeuSerGlyArgProThrProHisAlaArgAspIleAlaGlyIleSe 720

1957 .......CATTTAAACGACCATTGGTCGCAAAAAGAGGGCATTCCTCGCG 1999
720 rSerThrLysLysAspSerHisPheSerGlu..............AsnA 732

2000 GGGAAATCGTGTGGGACAACGACTGGATCAACCGCACATTTAAAGCGGAA 2049
```



```
3442 TTCCCC.....................................CGCGCCCG 3455
        |||                                     |||  ||
1213 ...ProThrValAsnSerGluSerSerAsnLysProLysAsnArgHisAr 1228

3456 CCGCGCCCGCCGGGATTTGCCG...CAACTGCAACCCCAACCGCAGCCCC 3502
     ||||:::  |||:::::::|||  :::::::::|||           :
1228 gArgSerValArgSerValProHisAsnValGluProAlaThrThrSerS 1245

3503 AACCGCAGCGC.................GACCTGATCAGCCGTTATGCC 3534
     ::  :::|||                 ||||||  |||     :::
1245 erAsnAspArgSerThrValAlaLeuCysAspLeuThrSerThrAsnThr 1261

3535 AATAGCGGTTTGAGTGAATTTTCCGCC..............ACGCTCAA 3569
     |||:::  ||||||:::     |||             :::|||||
1262 AsnAlaValLeuSerAspAlaArgAlaLysAlaGlnPheValAlaLeuAs 1278

3570 CAGCGTTTTCGCCGTACAGGACGAATTAGACCGCGTA...TTTGCCGAAG 3616
     |       ||||||:::::::::::::::::::::::   :::   ::::
1278 nValGlyLysAlaValSerGlnHisIleSerGlnLeuGluMetAsnAsnG 1295

3617 ACCGCCGCAACGCCGTTTGGACAAGCGGCATCCGGGACACCAAACACTAC 3666
     ::  :::     ||||||:::|||:::         :::|||:::|||
1295 luGlyGlnTyrAsnValTrpValSerAsnThrSerMetAsnLysAsnTyr 1311

3667 CGTTCGCAAGATTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCAAAT 3716
        |||::::::::::|||   :::   :::::::::        |||::
1312 SerSerSerGlnTyrArgArgPheSerSerLysSerThrGlnThrGlnLe 1328

3717 CGGTATGCAGAAAAACCTCGGCAGCGGG...CGCGTCGGCATCCTGTTTT 3763
     :|||   :::::::::::::::::::::   :::::::|||   :::|||:
1328 uGlyTrpAspGlnThrIleSerAsnAsnValGlnLeuGlyGlyValPheT 1345

3764 CGCACAACCGGACCGAAAACACCTTCGACGACGGCATCGGCAACTCGGCA 3813
     :::::   |||:::::::|||:::||||||   :::   ::::::::::::
1345 hrTyrValArgAsnSerAsnAsnPheAspLysAlaThrSerLysAsnThr 1361

3814 CGGCTTGCCCACGGCGCCGTTTTCGGGCAATACGGCATCGAC...AGGTT 3860
        ||||||:::          ::::::::::|||      |||   :::::
1362 ...LeuAlaGlnValAsnPheTyrSerLysTyrTyrAlaAspAsnHisTr 1377

3861 CTACATCGGCATCAGCGCGGGCGCGGGTTTTAGCAGCGGCAGCCTTTCAG 3910
     :|||:::||||||:::   |||   |||      ::::::::::|||:::
1377 pTyrLeuGlyIleAspLeuGlyTyrGlyLysPheGlnSerLysLeuGlnT 1394

3911 ACGGCATCGGAGGCAAAATCCGCCGCCGCGTGCTGCATTACGGCATTCAG 3960
       :::   ::::::|||:::   |||::::::   :::::::|||:::
1394 hrAsnHisAsnAlaLysPheAlaArgHisThrAlaGlnPheGlyLeuThr 1410

3961 GCACGATACCGCGCCGGTTTCGGCGGATTCGGCATCGAACCGCACATCGG 4010
     |||            ::::::|||:::|||||||||   |||   :::||
1411 AlaGlyLysAlaPheAsnLeuGlyAsnPheGlyIleThrProIleValGl 1427

4011 CGCAACGCGCTATTTCGTCCAAAAAGCGGATTACCGCTACGAAAACGTCA 4060
     |:::         ::::::::::::|||||:::      ::::::::::
1427 yValArgTyrSerTyrLeuSerAsnAlaAspPheAlaLeuAspGlnAlaA 1444

4061 ATATCGCCACCCCCGGCCTTGCATTCAACCGCTACCGCGCGGGCATTAAG 4110
     ::|||  :::     ::::::  :::         |||  :::
1444 rgIleLysValAsnProIleSerValLysThrAlaPheAlaGlnValAsp 1460

4111 GCAGATTATTCATTCAAACCGGCGCAACACATTTCCATCACGCCTTATTT 4160
        :::|||::::::      ::::::  :::|||:::||||||   ||
1461 LeuSerTyrThrTyrHisLeuGlyGlu...PheSerValThrProIleLe 1476

4161 GAGCCTGTCCTATACCGATGCCGCTTCGGGCAAAGTCCGAACACGCGTCA 4210
     ||||      |||   ||||||  :::|||:::  :::  :::||||
1476 uSerAlaArgTyr...AspAlaAsnGlnGlySerGlyLysIleAsnValA 1492

4211 ATACCGCCGTATTGGCTCAGGATTTCGGCAAAACCCGCAGTGCGGAATGG 4260
     ||         :::|||  :::      :::  ::::::  :::
1492 snGlyTyrAspPheAlaTyrAsnValGluAsnGlnGlnGlnTyrAsnAla 1508

4261 GGCGTAAACGCCGAAATCAAAGGTTTCACGCTGTCCCTCCACGCTGCCGC 4310
```

```
        |||::::::  :::      :::      |||||||||  ::::::
1509 GlyLeuLysLeuLysTyrHisAsnValLysLeuSerLeuIleGlyGlyLe 1525

4311 CGCCAAAGGCCCGCAACTGGAAGCGCAACACAGCGCGGGCATCAAATTAG 4360
      :::|||:::  |||   |||  |||  :::|||  :::|||||:
1525 uThrLysAlaLysGlnAlaGluLysGlnLysThrAlaGluLeuLysLeuS 1542

4361 GCTAC 4365
     :::::
1542 erPhe 1543
```

```
seq_name: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:PCT-US95-10661A-4

seq_documentation_block:
; Sequence 4, Application PC/TUS9510661A
;  GENERAL INFORMATION:
;    APPLICANT:  Washington University, et al.
;    TITLE OF INVENTION:  Haemophilus Adherence and Penetration Protein
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Flehr, Hohbach, Test, Albritton & Herbert
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/10661A
;      FILING DATE:  16-AUG-1995
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/296,791
;      FILING DATE:  25-AUG-1994
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Trecartin, Richard F.
;      REGISTRATION NUMBER:  31,801
;      REFERENCE/DOCKET NUMBER:  FP-59941/RFT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1545 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
PCT-US95-10661A-4

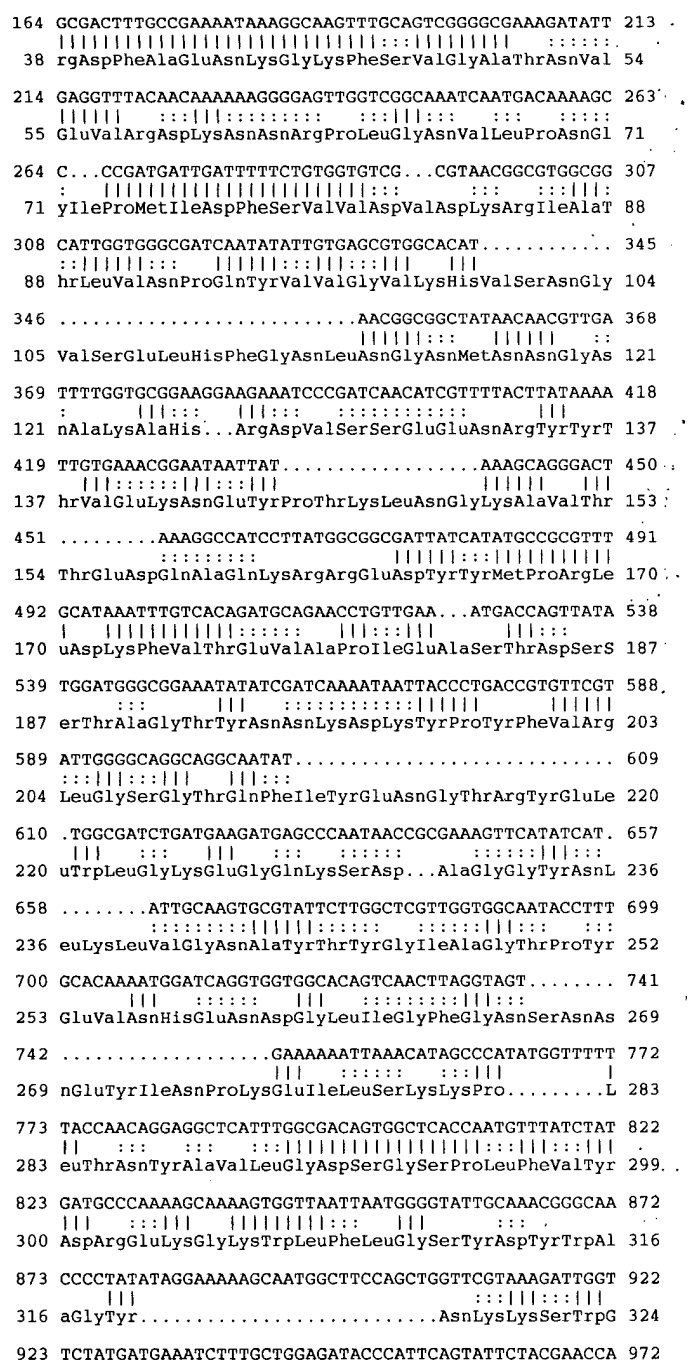

alignment_scores:
           Quality: 1579.50              Length:   1735
             Ratio:  1.704                Gaps:     53
 Percent Similarity:  53.429   Percent Identity: 26.686

alignment_block:
 US-09-303-518D-649 x PCT-US95-10661A-4   ..

 Align seg 1/1  to: PCT-US95-10661A-4  from: 1  to: 1545

      64 CGCTTCTCGCCTGCTTACTTAGCCATATGCCTGTCGTTCGGCATTCTTCC 113
         :::|||:::      :::::::|||:::   ::::::::::::::   ||
       5 LysPheLysLeuAsnPheIleAlaLeuThrValAlaTyrAlaLeuThrPr 21

     114 CCAAGCCTGGGCGGGACACACTTATTTCGGCATCAACTACCAATACTATC 163
         |   :::   |||:::   :::         :::::||||||   :::|
      21 oTyrThrGluAlaAlaLeuValArgAspAspValAspTyrGlnIlePheA 38
```


164 GCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATT 213
38 rgAspPheAlaGluAsnLysGlyLysPheSerValGlyAlaThrAsnVal 54
214 GAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAATCAATGACAAAAGC 263
55 GluValArgAspLysAsnAsnArgProLeuGlyAsnValLeuProAsnGl 71
264 C...CCGATGATTGATTTTTCTGTGGTGTCG...CGTAACGGCGTGGCGG 307
71 yIleProMetIleAspPheSerValValAspValAspLysArgIleAlaT 88
308 CATTGGTGGGCGATCAATATATTGTGAGCGTGGCACAT..... 345
88 hrLeuValAsnProGlnTyrValValGlyValLysHisValSerAsnGly 104
346 AACGGCGGCTATAACAACGTTGA 368
105 ValSerGluLeuHisPheGlyAsnLeuAsnGlyAsnMetAsnAsnGlyAs 121
369 TTTTGGTGCGGAAGGAAGAAATCCCGATCAACATCGTTTTACTTATAAAA 418
121 nAlaLysAlaHis...ArgAspValSerSerGluGluAsnArgTyrTyrT 137
419 TTGTGAAACGGAATAATTAT.....AAAGCAGGGACT 450
137 hrValGluLysAsnGluTyrProThrLysLeuAsnGlyLysAlaValThr 153
451 AAAGGCCATCCTTATGGCGGCGATTATCATATGCCGCGTTT 491
154 ThrGluAspGlnAlaGlnLysArgArgGluAspTyrTyrMetProArgLe 170
492 GCATAAATTTGTCACAGATGCAGAACCTGTTGAA...ATGACCAGTTATA 538
170 uAspLysPheValThrGluValAlaProIleGluAlaSerThrAspSerS 187
539 TGGATGGGCGGAAATATATCGATCAAAATAATTACCCTGACCGTGTTCGT 588
187 erThrAlaGlyThrTyrAsnAsnLysAspLysTyrProTyrPheValArg 203
589 ATTGGGGCAGGCAGGCAATAT..... 609
204 LeuGlySerGlyThrGlnPheIleTyrGluAsnGlyThrArgTyrGluLe 220
610 .TGGCGATCTGATGAAGATGAGCCCAATAACCGCGAAAGTTCATATCAT. 657
220 uTrpLeuGlyLysGluGlyGlnLysSerAsp...AlaGlyGlyTyrAsnL 236
658 ATTGCAAGTGCGTATTCTTGGCTCGTTGGTGGCAATACCTTT 699
236 euLysLeuValGlyAsnAlaTyrThrTyrGlyIleAlaGlyThrProTyr 252
700 GCACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGT..... 741
253 GluValAsnHisGluAsnAspGlyLeuIleGlyPheGlyAsnSerAsnAs 269
742 GAAAAAATTAAACATAGCCCATATGGTTTTT 772
269 nGluTyrIleAsnProLysGluIleLeuSerLysLysPro.....L 283
773 TACCAACAGGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATCTAT 822
283 euThrAsnTyrAlaValLeuGlyAspSerGlySerProLeuPheValTyr 299
823 GATGCCCAAAAGCAAAAGTGGTTAATTAATGGGGTATTGCAAACGGGCAA 872
300 AspArgGluLysGlyLysTrpLeuPheLeuGlySerTyrAspTyrTrpAl 316
873 CCCCTATATAGGAAAAAGCAATGGCTTCCAGCTGGTTCGTAAAGATTGGT 922
316 aGlyTyr.....AsnLysLysSerTrpG 324
923 TCTATGATGAAATCTTTGCTGGAGATACCCATTCAGTATTCTACGAACCA 972
324 lnGluTrpAsnIleTyrLysProGluPheAlaGluLysIleTyrGluGln 340
973 CGTCAAAATGGG.....AAATACTCTTTTAACGACGATAA 1007
341 TyrSerAlaGlySerLeuIleGlySerLysThrAspTyrSerTrpSerSe 357
1008 TAATGGCACAGGAAAAATCAATGCCAAACATGAACACAATTCTCTGCCTA 1057
357 rAsnGly..... 359
1058 ATAGATTAAAAACACGAACCGTT.....CAATTGTTTAATGTT 1095
360 LysThrSerThrIleThrGlyGlyGluLysSerLeuAsnVal 373
1096 TCTTTATCCGAGACAGCAAGAGAACCTGTTTATCATGCTGCAGGTGGTGT 1145
374 AspLeuAlaAspGlyLysAspLysPro..... 382
1146 CAACAGTTATCGACCCAGACTGAATAATGGAGAAAATATTTCCTTTATTG 1195
383 AsnHisGlyLysSerValThrPhe.... 390
1196 ACGAAGGAAAAGGCGAATTGATACTTACCAGCAACATCAATCAAGGTGCT 1245
391 ..GluGlySerGlyThrLeuThrLeuAsnAsnAsnIleAspGlnGlyAla 406
1246 GGAGGATTATATTTCCAAGGAGATTTTACGGTC...TCGCCTGAAAATAA 1292
407 GlyGlyLeuPhePheGluGlyAspTyrGluValLysGlyThrSerAspAs 423
1293 CGAAACTTGGCAAGGCGCGGGCGTTCATATCAGTGAAGACAGTACCGTTA 1342
423 nThrThrTrpLysGlyAlaGlyValSerValAlaGluGlyLysThrValT 440
1343 CTTGGAAAGTAAACGGCGTGGCAAACGACCGCCTGTCCAAAATCGGCAAA 1392
440 hrTrpLysValHisAsnProGlnTyrAspArgLeuAlaLysIleGlyLys 456
1393 GGCACGCTGCACGTTCAAGCCAAAGGGGAAAACCAAGGCTCGATCAGCGT 1442
457 GlyThrLeuIleValGluGlyThrGlyAspAsnLysGlySerLeuLysVa 473
1443 GGGCGACGGTACAGTCATTTTGGATCAGCAGGCAGACGATAAAGGCAAAA 1492
473 lGlyAspGlyThrValIleLeuLysGlnGlnThrAsnGlySerGly...G 489
1493 AACAAGCCTTTAGTGAAATCGGCTTGGTCAGCGGCAGGGGTACGGTGCAA 1542
489 lnHisAlaPheAlaSerValGlyIleValSerGlyArgSerThrLeuVal 505
1543 CTGAATGCCGATAATCAGTTCAACCCCGACAAACTCTATTTCGGCTTTCG 1592
506 LeuAsnAspAspLysGlnValAspProAsnSerIleTyrPheGlyPheAr 522
1593 CGGCGGACGTTTGGATTTAAACGGGCATTCGCTTTCGTTCCACCGTATTC 1642
522 gGlyGlyArgLeuAspLeuAsnGlyAsnSerLeuThrPheAspHisIleA 539
1643 AAAATACCGATGAAGGGGCGATGATTGTCAACCACAATCAAGACAAAGAA 1692
539 rgAsnIleAspGluGlyAlaArgLeuValAsnHisSerThrSerLysHis 555
1693 TCCACCGTTACCATTACAGGCAATAAAGATATTGCTACAACCGGCAATAA 1742
556 SerThrValThrIleThrGlyAsp...AsnLeuIleThrAspProAsnAs 571
1743 C..... 1743
571 nValSerIleTyrTyrValLysProLeuGluAspAspAsnProTyrAlaI 588
1743 1743

```
 588 leArgGlnIleLysTyrGlyTyrGlnLeuTyrPheAsnGluGluAsnArg 604

1743 ..................................................... 1743

 605 ThrTyrTyrAlaLeuLysLysAspAlaSerIleArgSerGluPheProGl 621

1743 ..................................................... 1743

 621 nAsnArgGlyGluSerAsnAsnSerTrpLeuTyrMetGlyThrGluLysA 638

1744 .....................AACAGCTTGGATAGCAAAAAAGAAATT 1770
                           |||   ::::::::::::::
 638 laAspAlaGlnLysAsnAlaMetAsnHisIleAsnAsnGluArgMetAsn 654

1771 GCCTACAACGGTTGGTTTGGCGAGAAAGATACGACCAAAACGAACGGGCG 1820
     ::::::||||||:::||||||||:::::::     |||:::||||||:: 
 655 GlyPheAsnGlyTyrPheGlyGluGluGlu...GlyLysAsnAsnGlyAs 670

1821 GCTCAACCTTGTTTACCAGCCCGCCGCAGAAGACCGCACCCTGCTGCTTT 1870
     :||||||:::::::::::::      :::|||:::::   :::|||||||: 
 670 nLeuAsnValThrPheLysGlyLysSerGluGlnAsnArgPheLeuLeuT 687

1871 CCGGCGGAACAAATTTAAACGGCAACATCACGCAAACAAACGGCAAACTG 1920
     ::|||||||||||||||||||||:::::::::      :::|||   ||| 
 687 hrGlyGlyThrAsnLeuAsnGlyAspLeuAsnValGlnGlnGlyThrLeu 703

1921 TTTTTCAGCGGCAGACCAACACCGCACGCCTACAAT..............  1956
     |||:::|||||||||||||||||||||||||   :::
 704 PheLeuSerGlyArgProThrProHisAlaArgAspIleAlaGlyIleSe 720

1957 .......CATTTAAACGACCATTGGTCGCAAAAAGAGGGCATTCCTCGCG 1999
               ::::::|||:::|||:::                ::::
 720 rSerThrLysLysAspSerHisPheSerGlu..............AsnA 732

2000 GGGAAATCGTGTGGGACAACGACTGGATCAACCGCACATTTAAAGCGGAA 2049
     ::|||:::|||   ::::::||||||||||||||||:::||||||||| 
 732 snGluValValValGluAspAspTrpIleAsnArgAsnPheLysAlaThr 748

2050 AACTTCCAAATTAAAGGCGGACAGGCGGTGGTTTCC...CGCAATGTTGC 2096
     |||:::::::::   ::::::   ::::::   |||   ||||||||| 
 749 AsnIleAsnValThrAsnAsnAlaThrLeuTyrSerGlyArgAsnValGl 765

2097 CAAAGTGAAAGGCGATTGGCATTTGAGCAATCACGCCCAAGCAGTTTTTG 2146
       ::::::   ::::::         ||||||:::|||::::::   :::| 
 765 uSerIleThrSerAsnIleThrAlaSerAsnAsnAlaLysValHisIleG 782

2147 GTGTCGCACCGCATCAAAGCCACACAATCTGTACACGTTCGGACTGGACG 2196
     ||                ::   |||:::|||:::|||||||||:::||| 
 782 lyTyr......LysAlaGlyAspThrValCysValArgSerAspTyrThr 796

2197 GGTCTGACAAATTGTGTCGAAAAAACCATTACCGACGATAAAGTGATTGC 2246
     |||   ::::::|||:::         ::::::   ||||||:::::: 
 797 GlyTyrValThrCysThrThrAspLysLeuSer...AspLysAlaLeuAs 812

2247 TTCATTGACTAAGACCGACATCAGCGGCAATGTCGATCTTGCCGATCACG 2296
      |||::::::   |||::::::   |||||||||:::|||::::::   | 
 812 nSerPheAsnProThrAsnLeuArgGlyAsnValAsnLeuThrGluSerA 829

2297 CTCATTTAAATCTCACAGGGCTTGCCACACTCAACGGCAATCTTAGTGCA 2346
     ||::::::   |||   |||   |||:::|||   |||:::::::::::: 
 829 laAsnPheValLeu...GlyLysAlaAsnLeuPheGlyThrIleGlnSer 844

2347 AATGGCGATACACGTTATACAGTCAGCCACAACGCCACCCAAAACGGCAA 2396
     :::|||:::::::::      :::::::::||| 
 845 ArgGlyAsnSerGlnValArgLeuThrGluAsn..................  855

2397 CCTTAGCCTCGTGGGCAATGCCCAAGCAACATTTAATCAAGCCACATTAA 2446

 855 ..................................................... 855

2447 ACGGCAACACATCGGCTTCGGGCAATGCTTCATTTAATCTAAGCGACCAC 2496

 855 ..................................................... 855
```

```
2497 GCCGTACAAAACGGCAGTCTGACGCTTTCCGGCAACGCTAAGGCAAACGT 2546

 855 ..................................................... 855

2547 AAGCCATTCCGCACTCAACGGTAATGTCTCCCTAGCCGATAAGGCAGTAT 2596

 855 ..................................................... 855

2597 TCCATTTTGAAAGCAGCCGCTTTACCGGACAAATCAGCGGCGGCAAGGAT 2646

 855 ..................................................... 855

2647 ACGGCATTACACTTAAAAGACAGCGAATGGACGCTGCCGTCAGGCACGGA 2696
                           |||:::|||   |||   :::::::::::
 856 .....................SerHisTrpHisLeuThrGlyAsnSerAs 865

2697 ATTAGGCAATTTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCT 2746
     ::::   :::|||:::|||   |||:::   |||   |||||||||||| 
 865 pValHisGlnLeuAspLeuAlaAsnGlyHisIleHisLeuAsnSerAlaA 882

2747 ATCGCCACGATGCGGCAGGGGCGCAAACCGGCAGTGCGACAGATGCGCCG 2796
       :::   :::
 882 spAsnSerAsn..................................... 885

2797 CGCCGCCGTTCGCGCCGTTCGCGCCGTTCCCTATTATCCGTTACACCGCC 2846
                                       :::||||||
 886 ...................................AsnValThr..... 888

2847 AACTTCGGTAGAATCCCGTTTCAACACGCTGACGGTAAACGGCAAATTGA 2896
                     ::::::||||||||||||||||||   :::|||: 
 889 ................LysTyrAsnThrLeuThrValAsn...SerLeuS 899

2897 ACGGTCAGGGAACATTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGC 2946
     ::|||:::|||:::|||   :::::::::::::|||   :::   :::::: 
 899 erGlyAsnGlySerPheTyrTyrLeuThrAspLeuSerAsnLysGlnGly 915

2947 GACAAATTGAAGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCGGT 2996
     ||||||:::   :::::::::|||:::   |||::::::||||||   || 
 916 AspLysValValValThrLysSerAlaThrGlyAsnPheThrLeuGlnVa 932

2997 CAACAATACCGGCAACGAACCTGCAAGCCTCGAACAATTGACGGTAGTGG 3046
     |   :::      :::||||||   :::   ::::::||||||:::   : 
 932 lAlaAspLysThrGlyGluPro...AsnHisAsnGluLeuThrLeuPheA 948

3047 AAGGAAAAGACAACAAACCGCTGTCCGAAAACCTTAATTTCACCCTGCAA 3096
     :::::      :::|||         ::::::||||||   :::||| 
 948 spAla......SerLysAlaGlnArgAspHisLeuAsnValSerLeuVal 962

3097 AACGAACACGTCGATGCCGGCGCGTGGCGTTACCAACTCATCCGCAAAGA 3146
     ::::::   ||||||   |||||||||:::|||:::|||   :::   :: 
 963 GlyAsnThrValAspLeuGlyAlaTrpLysTyrLysLeuArgAsnValAs 979

3147 CGGCGAGTTCCGCCTGCATAATCCGGTCAAAGAACAA..............  3183
     :|||::::::   |||:::||||||      |||:::
 979 nGlyArgTyrAspLeuTyrAsnProGluValGluLysArgAsnGlnThrV 996

3183 ..................................................... 3183

 996 alAspThrThrAsnIleThrThrProAsnAsnIleGlnAlaAspValPro 1012

3184 .........GAGCTTTCCGACAAACTCGGCAAG.................. 3207
               :::   :::::::::::::::::
1013 SerValProSerAsnAsnGluGluIleAlaArgValAspGluAlaProVa 1029

3208 .............................................GCAG 3211
                                                 ||||
1029 lProProProAlaProAlaThrProSerGluThrThrGluThrValAlaG 1046

3212 AAGCCAAAAAAACAGGCGGAAAAA.........GACAACGCGCAAAGCCTT 3252
     ||   :::||||||   :::|||            |||   |||::: 
1046 luAsnSerLysGlnGluSerLysThrValGluLysAsnGluGlnAspAla 1062
```

```
3253 GACGCGCTGATTGCGGCCGGGCGCGATGCCGTCGAA..... 3288
                 |||   :::|||:::::::::::::
1063 ThrGluThrThrAlaGlnAsnArgGluValAlaLysGluAlaLysSerAs 1079

3289 ....AAGACAGAAAGCGTTGCCGAACCGGCCCGGCAGGCAGGCGGGGAAA 3334
         |||:::::::::   ::::::   :::   |||:::|||:::|||:
1079 nValLysAlaAsnThrGlnThrAsnGluValAlaGlnSerGlySerGluT 1096

3335 ATGTCGGCATTATGCAGGCGGAG..................... 3357
     ::         :::   :::|||
1096 hrLysGluThrGlnThrThrGluThrLysGluThrAlaThrValGluLys 1112

3358 GAAGAGAAAAAACGGGTGCAGGCGGATAAAGACACCGCCTTGGCGAAA.. 3405
     |||||||||   :::|||:::::::::|||         :::   |||
1113 GluGluLysAlaLysValGluThrGluLysThrGlnGluValProLysVa 1129

3406 .....................................CAGCGCG 3412
                                          |||  :
1129 lThrSerGlnValSerProLysGlnGluGlnSerGluThrValGlnProG 1146

3413 AAGCGGAA..................................... 3420
     ::||||||
1146 lnAlaGluProAlaArgGluAsnAspProThrValAsnIleLysGluPro 1162

3420 ..... 3420

1163 GlnSerGlnThrAsnThrThrAlaAspThrGluGlnProAlaLysGluTh 1179

3420 ..... 3420

1179 rSerSerAsnValGluGlnProValThrGluSerThrThrValAsnThrG 1196

3421 ..........................ACCCGGCCGGCTACCACCGCC 3441
                                 |||   ||||||||||||
1196 lyAsnSerValValGluAsnProGluAsnThrThrProAlaThrThrGln 1212

3442 TTCCCC..............................CGCGCCCG 3455
        |||                              |||   ||
1213 ...ProThrValAsnSerGluSerSerAsnLysProLysAsnArgHisAr 1228

3456 CCGCGCCCGCCGGGATTTGCCG...CAACTGCAACCCCAACCGCAGCCCC 3502
     ||||:::   |||::::::|||   ::::::::::|||           :
1228 gArgSerValArgSerValProHisAsnValGluProAlaThrThrSerS 1245

3503 AACCGCAGCGC.................GACCTGATCAGCCGTTATGCC 3534
     ::   :::|||                 ||||||   |||      :::
1245 erAsnAspArgSerThrValAlaLeuCysAspLeuThrSerThrAsnThr 1261

3535 AATAGCGGTTTGAGTGAATTTTCCGCC..............ACGCTCAA 3569
     |||:::   ||||||:::      |||             :::|||||
1262 AsnAlaValLeuSerAspAlaArgAlaLysAlaGlnPheValAlaLeuAs 1278

3570 CAGCGTTTTCGCCGTACAGGACGAATTAGACCGCGTA...TTTGCCGAAG 3616
     |         ||||||:::::::::::::::::::::   :::   ::::
1278 nValGlyLysAlaValSerGlnHisIleSerGlnLeuGluMetAsnAsnG 1295

3617 ACCGCCGCAACGCCGTTTGGACAAGCGGCATCCGGGACACCAAACACTAC 3666
     ::   :::      ||||||:::|||:::         :::|||:::|||
1295 luGlyGlnTyrAsnValTrpValSerAsnThrSerMetAsnLysAsnTyr 1311

3667 CGTTCGCAAGATTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCAAAT 3716
        |||::::::::::|||   :::   :::::::::         |||::
1312 SerSerSerGlnTyrArgArgPheSerSerLysSerThrGlnThrGlnLe 1328

3717 CGGTATGCAGAAAAACCTCGGCAGCGGG...CGCGTCGGCATCCTGTTTT 3763
     :|||   :::::::::::::::::::::   ::::::|||   :::||||:
1328 uGlyTrpAspGlnThrIleSerAsnAsnValGlnLeuGlyGlyValPheT 1345

3764 CGCACAACCGGACCGAAAACACCTTCGACGACGGCATCGGCAACTCGGCA 3813
     :::::   |||::::::|||:::||||||   :::   ::::::::::::
1345 hrTyrValArgAsnSerAsnAsnPheAspLysAlaThrSerLysAsnThr 1361

3814 CGGCTTGCCCACGGCGCCGTTTTCGGGCAATACGGCATCGAC...AGGTT 3860
```

```
        ||||||:::         :::::::::|||      |||   :::::
1362 ...LeuAlaGlnValAsnPheTyrSerLysTyrTyrAlaAspAsnHisTr 1377

3861 CTACATCGGCATCAGCGCGGGCGCGGGTTTTAGCAGCGGCAGCCTTTCAG 3910
     :|||:::||||||:::   |||   |||      ::::::::::|||:::
1377 pTyrLeuGlyIleAspLeuGlyTyrGlyLysPheGlnSerLysLeuGlnT 1394

3911 ACGGCATCGGAGGCAAAATCCGCCGCCGCGTGCTGCATTACGGCATTCAG 3960
       :::   :::::::|||:::   |||:::::::   ::::::|||:::
1394 hrAsnHisAsnAlaLysPheAlaArgHisThrAlaGlnPheGlyLeuThr 1410

3961 GCACGATACCGCGCCGGTTTCGGCGGATTCGGCATCGAACCGCACATCGG 4010
     |||           ::::::|||:::|||||||||   |||   :::||
1411 AlaGlyLysAlaPheAsnLeuGlyAsnPheGlyIleThrProIleValGl 1427

4011 CGCAACGCGCTATTTCGTCCAAAAAGCGGATTACCGCTACGAAAACGTCA 4060
     |:::         ::::::::::::||||||:::      ::::::::::
1427 yValArgTyrSerTyrLeuSerAsnAlaAspPheAlaLeuAspGlnAlaA 1444

4061 ATATCGCCACCCCCGGCCTTGCATTCAACCGCTACCGCGCGGGCATTAAG 4110
     ::|||   :::      ::::::   :::         |||   :::
1444 rgIleLysValAsnProIleSerValLysThrAlaPheAlaGlnValAsp 1460

4111 GCAGATTATTCATTCAAACCGGCGCAACACATTTCCATCACGCCTTATTT 4160
        :::|||:::::      ::::::   :::|||:::||||||   ||
1461 LeuSerTyrThrTyrHisLeuGlyGlu...PheSerValThrProIleLe 1476

4161 GAGCCTGTCCTATACCGATGCCGCTTCGGGCAAAGTCCGAACACGCGTCA 4210
     ||||      |||   ||||||   :::|||:::   :::   :::||||
1476 uSerAlaArgTyr...AspAlaAsnGlnGlySerGlyLysIleAsnValA 1492

4211 ATACCGCCGTATTGGCTCAGGATTTCGGCAAAACCCGCAGTGCGGAATGG 4260
     ||         :::|||   :::      :::   ::::::   :::
1492 snGlyTyrAspPheAlaTyrAsnValGluAsnGlnGlnGlnTyrAsnAla 1508

4261 GGCGTAAACGCCGAAATCAAAGGTTTCACGCTGTCCCTCCACGCTGCCGC 4310
     |||:::::   :::      :::      |||||||||   ::::::
1509 GlyLeuLysLeuLysTyrHisAsnValLysLeuSerLeuIleGlyGlyLe 1525

4311 CGCCAAAGGCCCGCAACTGGAAGCGCAACACAGCGCGGGCATCAAATTAG 4360
      :::|||:::   |||   |||   |||   :::|||   :::||||||:
1525 uThrLysAlaLysGlnAlaGluLysGlnLysThrAlaGluLeuLysLeuS 1542

4361 GCTAC 4365
     :::::
1542 erPhe 1543
```

seq_name: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-296-791-5

```
seq_documentation_block:
; Sequence 5, Application US/08296791
; Patent No. 6245337
;  GENERAL INFORMATION:
;    APPLICANT:  St. Geme III, Joseph W.
;    APPLICANT:  Falkow, Stanley
;    TITLE OF INVENTION:  Haemophilus Adherence and Penetration
;    TITLE OF INVENTION:  Protein
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Flehr, Hohbach, Test, Albritton & Herbert
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/296,791
;      FILING DATE:  25-AUG-1994
```

```
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Trecartin, Richard F.
;      REGISTRATION NUMBER:  31,801
;      REFERENCE/DOCKET NUMBER:  A-59941/RFT/RMS
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1702 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
US-08-296-791-5


alignment_scores:
            Quality: 1568.00              Length:   1872
              Ratio:  1.673                Gaps:     49
 Percent Similarity:  50.053   Percent Identity: 24.947

alignment_block:
 US-09-303-518D-649 x US-08-296-791-5   ..

 Align seg 1/1  to: US-08-296-791-5  from: 1  to: 1702

      64 CGCTTCTCGCCTGCTTACTTAGCCATATGCCTGTCGTTCGGCATTCTTCC 113
         :::|||:::      :::::::|||:::   :::::::::::::::   ||
       5 LysPheLysLeuAsnPheIleAlaLeuThrValAlaTyrAlaLeuThrPr 21

     114 CCAAGCCTGGGCGGGACACACTTATTTCGGCATCAACTACCAATACTATC 163
         |   :::   |||:::   :::         ::::::||||||   :::|
      21 oTyrThrGluAlaAlaLeuValArgAspAspValAspTyrGlnIlePheA 38

     164 GCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATT 213
         ||||||||||||||||||||||||:::|||:::|||||||||   ::::::
      38 rgAspPheAlaGluAsnLysGlyArgPheSerValGlyAlaThrAsnVal 54

     214 GAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAATCAATGACAAAAGC 263
         ||||||   :::|||:::::::::   :::|||:::   :::   :::::
      55 GluValArgAspLysAsnAsnHisSerLeuGlyAsnValLeuProAsnGl 71

     264 C...CCGATGATTGATTTTTCTGTGGTGTCG...CGTAACGGCGTGGCGG 307
         :   |||||||||||||||||||||||:::      :::   :::|||:
      71 yIleProMetIleAspPheSerValValAspValAspLysArgIleAlaT 88

     308 CATTGGTGGGCGATCAATATATTGTGAGCGTGGCACAT...AACGGCGGC 354
         ::|||::::::   ||||||:::|||:::|||   |||   ::::::|||
      88 hrLeuIleAsnProGlnTyrValValGlyValLysHisValSerAsnGly 104

     355 TATAACAACGTTGATTTTGGT..............GCGGAAGGAAGAAA 389
         :::::::::   ||||||                  ::::::   ||
     105 ValSerGluLeuHisPheGlyAsnLeuAsnGlyAsnMetAsnAsnGlyAs 121

     390 TCCCGATCAACATCGTTTTACT..................TATAAAATTG 421
         |      :::||||||   :::                  |||      |
     121 nAspLysSerHisArgAspValSerSerGluGluAsnArgTyrPheServ 138

     422 TGAAACGGAATAATTATAAAGCAGGGACTAAAGGC.............. 456
         ||:::::|||:::|||   :::      :::|||
     138 alGluLysAsnGluTyrProThrLysLeuAsnGlyLysAlaValThrThr 154

     457 ............CATCCTTATGGCGGCGATTATCATATGCCGCGTTTGCA 494
                     :::          ||||||:::||||||||||||
     155 GluAspGlnThrGlnLysArgArgGluAspTyrTyrMetProArgLeuAs 171

     495 TAAATTTGTCACAGATGCAGAACCTGTTGAA...ATGACCAGTTATATGG 541
          ||||||||||||:::::   |||:::|||      |||:::      |
     171 pLysPheValThrGluValAlaProIleGluAlaSerThrAlaSerSerA 188

     542 ATGGGCGGAAATATATCGATCAAAATAATTACCCTGACCGTGTTCGTATT 591
         ||:::      |||   |||||||||:::||||||      ||||||:::

     188 spAlaGlyThrTyrAsnAspGlnAsnLysTyrProAlaPheValArgLeu 204

     592 GGGGCAGGCAGGCAATATTGGCGATCTGATGAAGATGAG........... 630
         |||:::|||   |||:::      :::      |||:::
     205 GlySerGlyThrGlnPheIleTyrLysLysGlyAspAsnTyrSerLeuIl 221

     631 .CCCAATAACCGCGAAAGTTCATATCAT............ATTGCAAGTG 667
             ||||||:::|||   :::   :::           :::::::::|
     221 eLeuAsnAsnHisGluValGlyGlyAsnAsnLeuLysLeuValGlyAspA 238

     668 CGTATTCTTGGCTCGTTGGTGGCAATACCTTTGCACAAAATGGATCAGGT 717
         |||||::::::   ::::::|||:::   :::      |||   ::::::
     238 laTyrThrTyrGlyIleAlaGlyThrProTyrLysValAsnHisGluAsn 254

     718 GGTGGCACAGTCAACTTAGGTAGTGAAAAAATTAAACATAGC...CCATA 764
         :::|||   ::::::::::|||:::::::|||   :::|||||   |||
     255 AsnGlyLeuIleGlyPheGlyAsnSerLysGluGluHisSerAspProLy 271

     765 TGGTTTTTTACCAACAGGA..............GGCTCATTTGGCGACA 799
          |||:::|||                      :::   :::|||||||
     271 sGlyIleLeuSerGlnAspProLeuThrAsnTyrAlaValLeuGlyAspS 288

     800 GTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGGTTAATT 849
         ||||||||||||:::|||:::||||||   :::|||   |||||||||:::
     288 erGlySerProLeuPheValTyrAspArgGluLysGlyLysTrpLeuPhe 304

     850 AATGGGGTATTGCAAACGGGCAACCCCTATATAGGAAAAAGCAATGGCTT 899
            |||      :::           |||      ||||||
     305 LeuGlySerTyrAspPheTrpAlaGlyTyrAsnLysLysSerTrpGlnGl 321

     900 CCAGCTGGTTCGTAAAGATTGGTTCTATGATGAAATCTTTGCTGGAGATA 949
             :::   |||      |||         ::::::      ||||
     321 uTrpAsnIleTyrLysProGluPheAlaLysThrValLeuAspLysAspT 338

     950 CCCATTCAGTATTCTACGAACCACGTCAAAATGGGAAATACTCTTTTAAC 999
         ||   :::   :::            :::|||   :::|||::::::|||
     338 hrAlaGlySerLeuIle......GlySerAsnThrGlnTyrAsnTrpAsn 352

    1000 GACGATAATAATGGCACAGGAAAAATCAATGCCAAACATGAACACAATTC 1049
                      |||||||||   ::::::
     353 Pro............ThrGlyLysThrSerVal..................  359

    1050 TCTGCCTAATAGATTAAAAACACGAACCGTTCAATTGTTTAATGTTTCTT 1099
         :::   |||      ::::::   :::|||:::||||||:::   |||
     360 .IleSerAsnGlySerGluSerLeuAsnValAspLeuPheAspSerSerG 376

    1100 TATCCGAGACAGCAAGAGAACCTGTTTATCATGCTGCAGGTGGTGTCAAC 1149
           :::      ::::::
     376 lnAspThrAspSerLys.............................. 381

    1150 AGTTATCGACCCAGACTGAATAATGGAGAAAATATTTCCTTTATTGACGA 1199
                    :::   |||:::|||:::::::::::::::::      ::
     382 ............LysAsnAsnHisGlyLysSerValThrLeu......Ar 392

    1200 AGGAAAAGGCGAATTGATACTTACCAGCAACATCAATCAAGGTGCTGGAG 1249
         :|||:::|||   |||   |||:::::|||||:::||||||||||||||
     392 gGlySerGlyThrLeuThrLeuAsnAsnAsnIleAspGlnGlyAlaGlyG 409

    1250 GATTATATTTCCAAGGAGATTTTACGGTC...TCGCCTGAAAATAACGAA 1296
         ||||:::|||:::||||||:::   |||   :::   :::::::::
     409 lyLeuPhePheGluGlyAspTyrGluValLysGlyThrSerAspSerThr 425

    1297 ACTTGGCAAGGCGCGGGCGTTCATATCAGTGAAGACAGTACCGTTACTTG 1346
         ||||||:::||||||||||||   :::::::::   :::||||||||||||
     426 ThrTrpLysGlyAlaGlyValSerValAlaAspGlyLysThrValThrTr 442

    1347 GAAAGTAAACGGCGTGGCAAACGACCGCCTGTCCAAAATCGGCAAAGGCA 1396
         |||||||:::::      :::||||||||||:::||||||||||||||
     442 pLysValHisAsnProLysSerAspArgLeuAlaLysIleGlyLysGlyT 459

    1397 CGCTGCACGTTCAAGCCAAAGGGGAAAACCAAGGCTCGATCAGCGTGGGC 1446
         |||||   |||:::::|||||||||||||:::||||||:::::|||||||
     459 hrLeuIleValGluGlyLysGlyGluAsnLysGlySerLeuLysValGly 475
```



```
                ::::|||||||:::::::   ::::::|||||||:::   ::::::
 937 LysThrGlyGluProThrLys...AsnGluLeuThrLeuPheAspAla.. 951

3054 AGACAACAAACCGCTGTCCGAAAACCTTAATTTCACCCTGCAAAACGAAC 3103
       ::::::          :::|||||||||   :::|||   ::::::
 952 ....SerAsnAlaThrArgAsnAsnLeuAsnValSerLeuValGlyAsnT 967

3104 ACGTCGATGCCGGCGCGTGGCGTTACCAACTCATCCGCAAAGACGGCGAG 3153
       ||||||   |||||||||:::|||:::|||   :::   :::|||:::
 967 hrValAspLeuGlyAlaTrpLysTyrLysLeuArgAsnValAsnGlyArg 983

3154 TTCCGCCTGCATAATCCG..................................... 3171
     :::   |||:::||||||
 984 TyrAspLeuTyrAsnProGluValGluLysArgAsnGlnThrValAspTh 1000

3171 ..... 3171

1000 rThrAsnIleThrThrProAsnAsnIleGlnAlaAspValProSerValP 1017

3171 ..... 3171

1017 roSerAsnAsnGluGluIleAlaArgValGluThrProValProProPro 1033

3171 ..... 3171

1034 AlaProAlaThrProSerGluThrThrGluThrValAlaGluAsnSerLy 1050

3171 ..... 3171

1050 sGlnGluSerLysThrValGluLysAsnGluGlnAspAlaThrGluThrT 1067

3171 ..... 3171

1067 hrAlaGlnAsnGlyGluValAlaGluGluAlaLysProSerValLysAla 1083

3171 ..... 3171

1084 AsnThrGlnThrAsnGluValAlaGlnSerGlySerGluThrGluGluTh 1100

3172 .............GTCAAAGAACAAGAGCTTTCCGACAAACTCGGCAAGG 3208
                 :::||||||            :::|||   ||||
1100 rGlnThrThrGluIleLysGluThrAlaLysValGluLysGluGluLysA 1117

3209 CAGAAGCCAAAAAACAG..................................... 3225
     ||:::::::::|||:::
1117 laLysValGluLysGluGluLysAlaLysValGluLysAspGluIleGln 1133

3225 ..... 3225

1134 GluAlaProGlnMetAlaSerGluThrSerProLysGlnAlaLysProAl 1150

3225 ..... 3225

1150 aProLysGluValSerThrAspThrLysValGluGluThrGlnValGlnA 1167

3225 ..... 3225

1167 laGlnProGlnThrGlnSerThrThrValAlaAlaAlaGluAlaThrSer 1183

3226 ..............................GCGGAAAAAGACAACGC 3242
                                   :::|||||   |||||
1184 ProAsnSerLysProAlaGluGluThrGlnProSerGluLysThrAsnAl 1200

3243 GCAAAGCCTTGACGCGCTGATTGCG..................... 3267
     |:::   :::      :::::::::
1200 aGluProValThrProValValSerLysAsnGlnThrGluAsnThrThrA 1217

3268 ......................GCCGGGCGCGATGCCGTCGAAAAGACA 3294
                            :::::::::     :::|||||||||
1217 spGlnProThrGluArgGluLysThrAlaLysValGluThrGluLysThr 1233

3295 GAAAGCGTTGCCGAACCGGCCCGGCAGGCAGGC.................. 3327
     ::::::      :::   |||   |||||||:::
```

```
1234 GlnGluProProGlnValAlaSerGlnAlaSerProLysGlnGluGlnSe 1250

3328 ..............................GGGGAAAATGTC...GGCA 3343
                                    :::|||||||||      :
1250 rGluThrValGlnProGlnAlaValLeuGluSerGluAsnValProThrV 1267

3344 TTATGCAGGCGGAGGAAGAGAAAAAACGGGTGCAGGCGGATAAAGAC... 3390
     ::   :::|||||||||   :::   :::::::|||::::::   :::
1267 alAsnAsnAlaGluGluValGlnAlaGlnLeuGlnThrGlnThrSerAla 1283

3390 ..... 3390

1284 ThrValSerThrLysGlnProAlaProGluAsnSerIleAsnThrGlySe 1300

3391 ....ACCGCCTTGGCGAAACAGCGCGAAGCGGAAACCCGGCCGGCTACC. 3435
       ||||||:::::::::      |||   :::  :::|||   |||
1300 rAlaThrAlaIleThrGluThrAlaGluLysSerAspLysProGlnThrG 1317

3435 ..... 3435

1317 luThrAlaAlaSerThrGluAspAlaSerGlnHisLysAlaAsnThrVal 1333

3436 ..............................ACCGCCTTCCCCCGCGCCCG 3455
                                      ::::::   |||:::::::||
1334 AlaAspAsnSerValAlaAsnAsnSerGluSerSerGluProLysSerAr 1350

3456 CCGCGCCCGCCGGGATTTGCCGCAACTGCAA..................... 3486
     ||||   ||||||::::::   |||   |||
1350 gArg...ArgArgSerIleSerGlnProGlnGluThrSerAlaGluGluT 1366

3486 ..... 3486

1366 hrThrAlaAlaSerThrAspGluThrThrIleAlaAspAsnSerLysArg 1382

3487 CCCCAACCGCAGCCCCAACCGCAG.......................... 3510
       :::|||:::   :::   :::
1383 SerLysProAsnArgArgSerArgArgSerValArgSerGluProThrVa 1399

3511 ..............................CGCGACCTGATCAGCC 3526
                                    |||||||||   |||
1399 lThrAsnGlySerAspArgSerThrValAlaLeuArgAspLeuThrSerT 1416

3527 GTTATGCCAATAGCGGTTTGAGTGAATTTTCCGCC............... 3561
       :::|||:::   :::|||:::      |||
1416 hrAsnThrAsnAlaValIleSerAspAlaMetAlaLysAlaGlnPheVal 1432

3562 ACGCTCAACAGCGTTTTCGCCGTACAGGACGAATTAGACCGCGTA...TT 3608
     :::||||||         ||||||:::::::::::::::::::::::   ::
1433 AlaLeuAsnValGlyLysAlaValSerGlnHisIleSerGlnLeuGluMe 1449

3609 TGCCGAAGACCGCCGCAACGCCGTTTGGACAAGCGGCATCCGGGACACCA 3658
     :   ::::::   :::       ||||||:::|||:::          ::::
1449 tAsnAsnGluGlyGlnTyrAsnValTrpValSerAsnThrSerMetAsnG 1466

3659 AACACTACCGTTCGCAAGATTTCCGCGCCTACCGCCAACAAACCGACCTG 3708
     ::::::|||   |||::::::::::|||   :::   :::::::::
1466 luAsnTyrSerSerSerGlnTyrArgArgPheSerSerLysSerThrGln 1482

3709 CGCCAAATCGGTATGCAGAAAAACCTCGGCAGCGGG...CGCGTCGGCAT 3755
       |||:::|||   ::::::::::::::::::::::   :::::::|||
1483 ThrGlnLeuGlyTrpAspGlnThrIleSerAsnAsnValGlnLeuGlyGl 1499

3756 CCTGTTTTCGCACAACCGGACCGAAAACACCTTCGACGACGGCATCGGCA 3805
     :::|||::::::   |||:::::|||:::||||||   :::   ::::
1499 yValPheThrTyrValArgAsnSerAsnAsnPheAspLysAlaSerSerL 1516

3806 ACTCGGCACGGCTTGCCCACGGCGCCGTTTTCGGGCAATACGGCATCGAC 3855
     ::::::::   ||||||:::          :::::::::|||      |||
1516 ysAsnThr...LeuAlaGlnValAsnPheTyrSerLysTyrTyrAlaAsp 1531

3856 ...AGGTTCTACATCGGCATCAGCGCGGGCGCGGGTTTTAGCAGCGGCAG 3902
       ::::::|||:::||||||:::   |||   |||      ::::::::
1532 AsnHisTrpTyrLeuGlyIleAspLeuGlyTyrGlyLysPheGlnSerAs 1548
```

```
    3903 CCTTTCAGACGGCATCGGAGGCAAAATCCGCCCGCCGCGTGCTGCATTACG 3952
         :|||:::   :::   ::::::|||:::   |||::::::   ::::::|
    1548 nLeuLysThrAsnHisAsnAlaLysPheAlaArgHisThrAlaGlnPheG 1565

    3953 GCATTCAGGCACGATACCGCGCCGGTTTCGGCGGATTCGGCATCGAACCG 4002
         ||:::   |||           :::::::|||:::||||||||||   |||
    1565 lyLeuThrAlaGlyLysAlaPheAsnLeuGlyAsnPheGlyIleThrPro 1581

    4003 CACATCGGCGCAACGCGCTATTTCGTCCAAAAAGCGGATTACCGCTACGA 4052
            :::|||:::         ::::::::::|||::::::
    1582 IleValGlyValArgTyrSerTyrLeuSerAsnAlaAsnPheAlaLeuAl 1598

    4053 AAACGTCAATATCGCCACCCCCGGCCTTGCATTCAACCGCTACCGCGCGG 4102
          :::   :::|||   :::      ::::::   :::         |||
    1598 aLysAspArgIleLysValAsnProIleSerValLysThrAlaPheAlaG 1615

    4103 GCATTAAGGCAGATTATTCATTCAAACCGGCGCAACACATTTCCATCACG 4152
           :::      :::|||::::::      ::::::   :::|||:::|||
    1615 lnValAspLeuSerTyrThrTyrHisLeuGlyGlu...PheSerValThr 1630

    4153 CCTTATTTGAGCCTGTCCTAT...ACCGATGCCGCTTCGGGCAAAGTCCG 4199
         |||   ||||||      |||   |||:::   :::||||||||:::
    1631 ProIleLeuSerAlaArgTyrAspThrAsnGlnGlySerGlyLysIle.. 1646

    4200 AACACGCGTCAATACCGCCGTATTGGCTCAGGATTTCGGCAAAACCCGCA 4249
             :::||||||         :::|||   :::      :::   ::::
    1647 ....AsnValAsnGlnTyrAspPheAlaTyrAsnValGluAsnGlnGlnG 1662

    4250 GTGCGGAATGGGGCGTAAACGCCGAAATCAAAGGTTTCACGCTGTCCCTC 4299
         ::   :::   |||::::::   :::      :::      |||||||||
    1662 lnTyrAsnAlaGlyLeuLysLeuLysTyrHisAsnValLysLeuSerLeu 1678

    4300 CACGCTGCCGCCGCCAAAGGCCCGCAACTGGAAGCGCAACACAGCGCGGG 4349
           ::::::   :::|||:::   |||   |||   |||   :::|||
    1679 IleGlyGlyLeuThrLysAlaLysGlnAlaGluLysGlnLysThrAlaGl 1695

    4350 CATCAAATTAGGCTAC 4365
          :::||||||::::::
    1695 uLeuLysLeuSerPhe 1700

seq_name: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:PCT-US95-10661A-5

seq_documentation_block:
; Sequence 5, Application PC/TUS9510661A
;  GENERAL INFORMATION:
;     APPLICANT:  Washington University, et al.
;     TITLE OF INVENTION:  Haemophilus Adherence and Penetration Protein
;     NUMBER OF SEQUENCES:  9
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Flehr, Hohbach, Test, Albritton & Herbert
;       STREET:  4 Embarcadero Center, Suite 3400
;       CITY:  San Francisco
;       STATE:  California
;       COUNTRY:  United States
;       ZIP:  94111-4187
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US95/10661A
;       FILING DATE:  16-AUG-1995
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/296,791
;       FILING DATE:  25-AUG-1994
;       CLASSIFICATION:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Trecartin, Richard F.
;       REGISTRATION NUMBER:  31,801
;       REFERENCE/DOCKET NUMBER:  FP-59941/RFT
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (415) 781-1989
;       TELEFAX:  (415) 398-3249
;       TELEX:  910 277299
;   INFORMATION FOR SEQ ID NO:  5:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1702 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  unknown
PCT-US95-10661A-5


alignment_scores:
            Quality: 1568.00              Length:   1872
              Ratio:  1.673                 Gaps:     49
 Percent Similarity:  50.053   Percent Identity: 24.947

alignment_block:
 US-09-303-518D-649 x PCT-US95-10661A-5   ..

 Align seg 1/1  to: PCT-US95-10661A-5  from: 1  to: 1702

      64 CGCTTCTCGCCTGCTTACTTAGCCATATGCCTGTCGTTCGGCATTCTTCC 113
         :::|||:::      ::::::|||:::   ::::::::::::::   ||
       5 LysPheLysLeuAsnPheIleAlaLeuThrValAlaTyrAlaLeuThrPr 21

     114 CCAAGCCTGGGCGGGACACACTTATTTCGGCATCAACTACCAATACTATC 163
         |   :::   |||:::   :::          ::::::||||||   :::|
      21 oTyrThrGluAlaAlaLeuValArgAspAspValAspTyrGlnIlePheA 38

     164 GCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATT 213
         ||||||||||||||||||||||||:::|||:::|||||||||   ::::::
      38 rgAspPheAlaGluAsnLysGlyArgPheSerValGlyAlaThrAsnVal 54

     214 GAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAATCAATGACAAAAGC 263
         ||||||   :::|||:::::::::   :::|||:::   :::   :::::
      55 GluValArgAspLysAsnAsnHisSerLeuGlyAsnValLeuProAsnGl 71

     264 C...CCGATGATTGATTTTTCTGTGGTGTCG...CGTAACGGCGTGGCGG 307
         :   ||||||||||||||||||||||||:::      :::   :::|||:
      71 yIleProMetIleAspPheSerValValAspValAspLysArgIleAlaT 88

     308 CATTGGTGGGCGATCAATATATTGTGAGCGTGGCACAT...AACGGCGGC 354
         ::|||::::::   ||||||:::|||:::|||   |||   ::::::|||
      88 hrLeuIleAsnProGlnTyrValValGlyValLysHisValSerAsnGly 104

     355 TATAACAACGTTGATTTTGGT..............GCGGAAGGAAGAAA 389
            :::::::::   ||||||                ::::::   ||
     105 ValSerGluLeuHisPheGlyAsnLeuAsnGlyAsnMetAsnAsnGlyAs 121

     390 TCCCGATCAACATCGTTTTACT..................TATAAAATTG 421
         |      :::||||||   :::                |||      |
     121 nAspLysSerHisArgAspValSerSerGluGluAsnArgTyrPheSerV 138

     422 TGAAACGGAATAATTATAAAGCAGGGACTAAAGGC.............. 456
         ||:::::::|||:::|||   :::      :::|||
     138 alGluLysAsnGluTyrProThrLysLeuAsnGlyLysAlaValThrThr 154

     457 ............CATCCTTATGGCGGCGATTATCATATGCCGCGTTTGCA 494
                     :::              ||||||:::||||||||||||
     155 GluAspGlnThrGlnLysArgArgGluAspTyrTyrMetProArgLeuAs 171

     495 TAAATTTGTCACAGATGCAGAACCTGTTGAA...ATGACCAGTTATATGG 541
          ||||||||||||::::::   |||:::|||      |||:::      |
     171 pLysPheValThrGluValAlaProIleGluAlaSerThrAlaSerSerA 188

     542 ATGGGCGGAAATATATCGATCAAAATAATTACCCTGACCGTGTTCGTATT 591
         ||:::      |||   |||||||||:::||||||      ||||||:::
     188 spAlaGlyThrTyrAsnAspGlnAsnLysTyrProAlaPheValArgLeu 204

     592 GGGGCAGGCAGGCAATATTGGCGATCTGATGAAGATGAG........... 630
         |||:::|||   |||:::      :::      |||:::
     205 GlySerGlyThrGlnPheIleTyrLysLysGlyAspAsnTyrSerLeuIl 221
```

```
 631 .CCCAATAACCGCGAAAGTTCATATCAT.....ATTGCAAGTG 667
 221 eLeuAsnAsnHisGluValGlyGlyAsnAsnLeuLysLeuValGlyAspA 238

 668 CGTATTCTTGGCTCGTTGGTGGCAATACCTTTGCACAAAATGGATCAGGT 717
 238 laTyrThrTyrGlyIleAlaGlyThrProTyrLysValAsnHisGluAsn 254

 718 GGTGGCACAGTCAACTTAGGTAGTGAAAAAATTAAACATAGC...CCATA 764
 255 AsnGlyLeuIleGlyPheGlyAsnSerLysGluGluHisSerAspProLy 271

 765 TGGTTTTTTACCAACAGGA.............GGCTCATTTGGCGACA 799
 271 sGlyIleLeuSerGlnAspProLeuThrAsnTyrAlaValLeuGlyAspS 288

 800 GTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGGTTAATT 849
 288 erGlySerProLeuPheValTyrAspArgGluLysGlyLysTrpLeuPhe 304

 850 AATGGGGTATTGCAAACGGGCAACCCCTATATAGGAAAAAGCAATGGCTT 899
 305 LeuGlySerTyrAspPheTrpAlaGlyTyrAsnLysLysSerTrpGlnGl 321

 900 CCAGCTGGTTCGTAAAGATTGGTTCTATGATGAAATCTTTGCTGGAGATA 949
 321 uTrpAsnIleTyrLysProGluPheAlaLysThrValLeuAspLysAspT 338

 950 CCCATTCAGTATTCTACGAACCACGTCAAAATGGGAAATACTCTTTTAAC 999
 338 hrAlaGlySerLeuIle......GlySerAsnThrGlnTyrAsnTrpAsn 352

1000 GACGATAATAATGGCACAGGAAAAATCAATGCCAAACATGAACACAATTC 1049
 353 Pro............ThrGlyLysThrSerVal................. 359

1050 TCTGCCTAATAGATTAAAAACACGAACCGTTCAATTGTTTAATGTTTCTT 1099
 360 .IleSerAsnGlySerGluSerLeuAsnValAspLeuPheAspSerSerG 376

1100 TATCCGAGACAGCAAGAGAACCTGTTTATCATGCTGCAGGTGGTGTCAAC 1149
 376 lnAspThrAspSerLys............................... 381

1150 AGTTATCGACCCAGACTGAATAATGGAGAAAATATTTCCTTTATTGACGA 1199
 382 ............LysAsnAsnHisGlyLysSerValThrLeu......Ar 392

1200 AGGAAAAGGCGAATTGATACTTACCAGCAACATCAATCAAGGTGCTGGAG 1249
 392 gGlySerGlyThrLeuThrLeuAsnAsnAsnIleAspGlnGlyAlaGlyG 409

1250 GATTATATTTCCAAGGAGATTTTACGGTC...TCGCCTGAAAATAACGAA 1296
 409 lyLeuPhePheGluGlyAspTyrGluValLysGlyThrSerAspSerThr 425

1297 ACTTGGCAAGGCGCGGGCGTTCATATCAGTGAAGACAGTACCGTTACTTG 1346
 426 ThrTrpLysGlyAlaGlyValSerValAlaAspGlyLysThrValThrTr 442

1347 GAAAGTAAACGGCGTGGCAAACGACCGCCTGTCCAAAATCGGCAAAGGCA 1396
 442 pLysValHisAsnProLysSerAspArgLeuAlaLysIleGlyLysGlyT 459

1397 CGCTGCACGTTCAAGCCAAAGGGGAAAACCAAGGCTCGATCAGCGTGGGC 1446
 459 hrLeuIleValGluGlyLysGlyGluAsnLysGlySerLeuLysValGly 475

1447 GACGGTACAGTCATTTTGGATCAGCAGGCAGACGATAAAGGCAAAAAACA 1496
 476 AspGlyThrValIleLeuLysGlnGlnAlaAspAlaAsnAsnLysValLy 492

1497 AGCCTTTAGTGAAATCGGCTTGGTCAGCGGCAGGGGTACGGTGCAACTGA 1546
 492 sAlaPheSerGlnValGlyIleValSerGlyArgSerThrValValLeuA 509

1547 ATGCCGATAATCAGTTCAACCCCGACAAACTCTATTTCGGCTTTCGCGGC 1596
 509 snAspAspLysGlnValAspProAsnSerIleTyrPheGlyPheArgGly 525

1597 GGACGTTTGGATTTAAACGGGCATTCGCTTTCGTTCCACCGTATTCAAAA 1646
 526 GlyArgLeuAspAlaAsnGlyAsnAsnLeuThrPheGluHisIleArgAs 542

1647 TACCGATGAAGGGGCGATGATTGTCAACCACAATCAAGACAAAGAATCCA 1696
 542 nIleAspAspGlyAlaArgLeuValAsnHisAsnThrSerLysThrSerT 559

1697 CCGTTACCATTACAGGC................................ 1713
 559 hrValThrIleThrGlyGluSerLeuIleThrAspProAsnThrIleThr 575

1713 ..................................................... 1713
 576 ProTyrAsnIleAspAlaProAspGluAspAsnProTyrAlaPheArgAr 592

1713 ..................................................... 1713
 592 gIleLysAspGlyGlyGlnLeuTyrLeuAsnLeuGluAsnTyrThrTyrT 609

1713 ..................................................... 1713
 609 yrAlaLeuArgLysGlyAlaSerThrArgSerGluLeuProLysAsnSer 625

1714 ..................................AATAAAGATATTGC 1727
 626 GlyGluSerAsnGluAsnTrpLeuTyrMetGlyLysThrSerAspAlaAl 642

1728 TACAACCGGCAATAACAACAGCTTGGATAGCAAAAAAGAAATTGCCTACA 1777
 642 aLysArgAsnValMetAsnHisIleAsnAsnGluArgMetAsnGlyPheA 659

1778 ACGGTTGGTTTGGCGAGAAAGATACGACCAAAACGAACGGGCGGCTCAAC 1827
 659 snGlyTyrPheGlyGluGluGlu...GlyLysAsnAsnGlyAsnLeuAsn 674

1828 CTTGTTTACCAGCCCGCCGCAGAAGACCGCACCCTGCTGCTTTCCGGCGG 1877
 675 ValThrPheLysGlyLysSerGluGlnAsnArgPheLeuLeuThrGlyGl 691

1878 AACAAATTTAAACGGCAACATCACGCAAACAAACGGCAAACTGTTTTTCA 1927
 691 yThrAsnLeuAsnGlyAspLeuLysValGluLysGlyThrLeuPheLeuS 708

1928 GCGGCAGACCAACACCGCACGCCTACAAT..................... 1956
 708 erGlyArgProThrProHisAlaArgAspIleAlaGlyIleSerSerThr 724

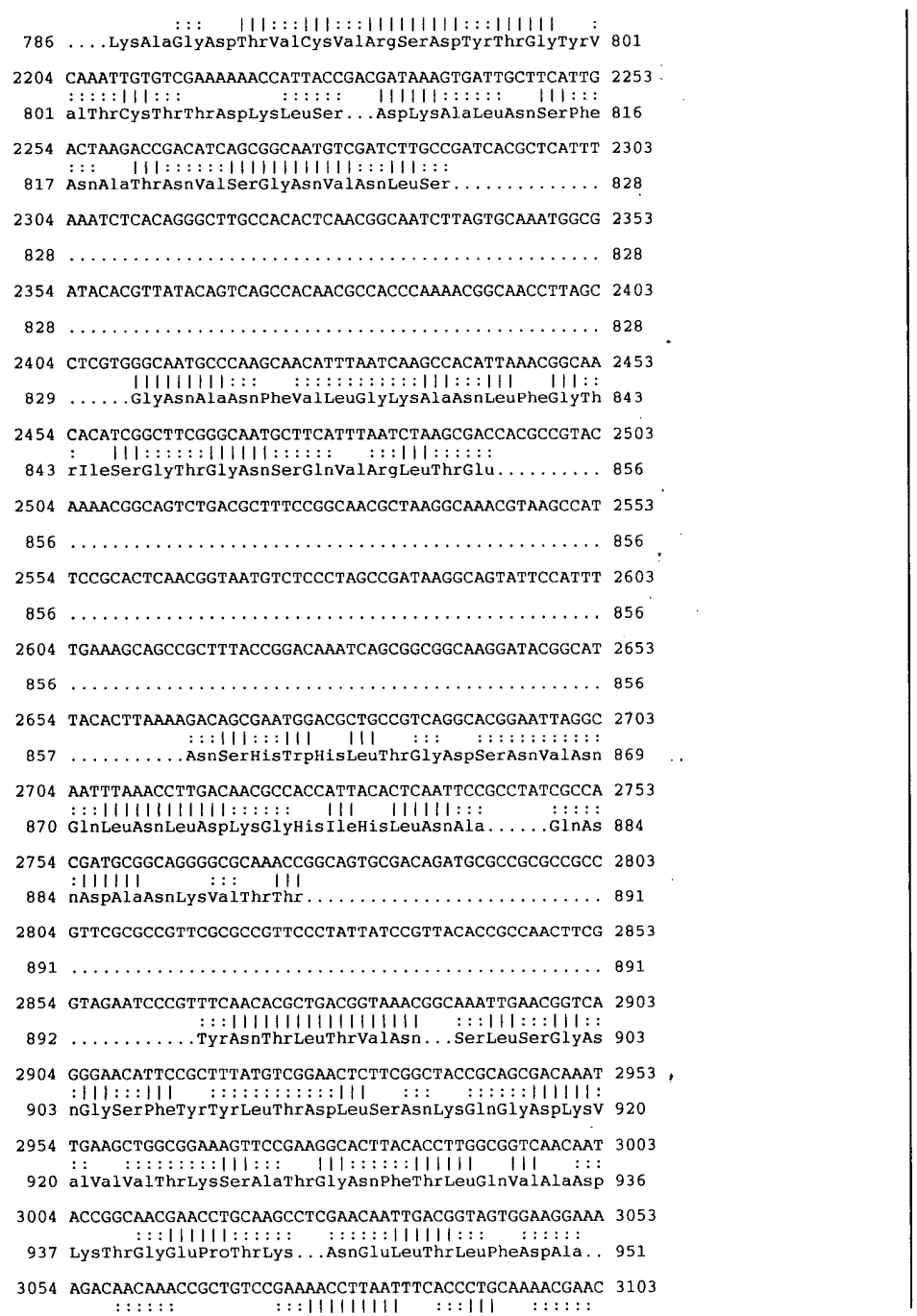
1957 CATTTAAACGACCATTGGTCGCAAAAAGAGGGCATTCCTCGCGGGGAAAT 2006
 725 LysLysAspGlnHisPheAlaGlu..............AsnAsnGluVa 736

2007 CGTGTGGGACAACGACTGGATCAACCGCACATTTAAAGCGGAAAACTTCC 2056
 736 lValValGluAspAspTrpIleAsnArgAsnPheLysAlaThrAsnIleA 753

2057 AAATTAAAGGCGGACAGGCGGTGGTTTCC...CGCAATGTTGCCAAAGTG 2103
 753 snValThrAsnAsnAlaThrLeuTyrSerGlyArgAsnValAlaAsnIle 769

2104 AAAGGCGATTGGCATTTGAGCAATCACGCCCAAGCAGTTTTTGGTGTCGC 2153
 770 ThrSerAsnIleThrAlaSerAspAsnAlaLysValHisIleGlyTyr.. 785

2154 ACCGCATCAAAGCCACACAATCTGTACACGTTCGGACTGGACGGGTCTGA 2203
```


786 LysAlaGlyAspThrValCysValArgSerAspTyrThrGlyTyrV 801
2204 CAAATTGTGTCGAAAAAACCATTACCGACGATAAAGTGATTGCTTCATTG 2253
801 alThrCysThrThrAspLysLeuSer...AspLysAlaLeuAsnSerPhe 816
2254 ACTAAGACCGACATCAGCGGCAATGTCGATCTTGCCGATCACGCTCATTT 2303
817 AsnAlaThrAsnValSerGlyAsnValAsnLeuSer..... 828
2304 AAATCTCACAGGGCTTGCCACACTCAACGGCAATCTTAGTGCAAATGGCG 2353
828 828
2354 ATACACGTTATACAGTCAGCCACAACGCCACCCAAAACGGCAACCTTAGC 2403
828 828
2404 CTCGTGGGCAATGCCCAAGCAACATTTAATCAAGCCACATTAAACGGCAA 2453
829 GlyAsnAlaAsnPheValLeuGlyLysAlaAsnLeuPheGlyTh 843
2454 CACATCGGCTTCGGGCAATGCTTCATTTAATCTAAGCGACCACGCCGTAC 2503
843 rIleSerGlyThrGlyAsnSerGlnValArgLeuThrGlu..... 856
2504 AAAACGGCAGTCTGACGCTTTCCGGCAACGCTAAGGCAAACGTAAGCCAT 2553
856 856
2554 TCCGCACTCAACGGTAATGTCTCCCTAGCCGATAAGGCAGTATTCCATTT 2603
856 856
2604 TGAAAGCAGCCGCTTTACCGGACAAATCAGCGGCGGCAAGGATACGGCAT 2653
856 856
2654 TACACTTAAAAGACAGCGAATGGACGCTGCCGTCAGGCACGGAATTAGGC 2703
857 AsnSerHisTrpHisLeuThrGlyAspSerAsnValAsn 869
2704 AATTTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCGCCA 2753
870 GlnLeuAsnLeuAspLysGlyHisIleHisLeuAsnAla.....GlnAs 884
2754 CGATGCGGCAGGGGCGCAAACCGGCAGTGCGACAGATGCGCCGCGCCGCC 2803
884 nAspAlaAsnLysValThrThr..... 891
2804 GTTCGCGCCGTTCGCGCCGTTCCCTATTATCCGTTACACCGCCAACTTCG 2853
891 891
2854 GTAGAATCCCGTTTCAACACGCTGACGGTAAACGGCAAATTGAACGGTCA 2903
892 TyrAsnThrLeuThrValAsn...SerLeuSerGlyAs 903
2904 GGGAACATTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGACAAAT 2953
903 nGlySerPheTyrTyrLeuThrAspLeuSerAsnLysGlnGlyAspLysV 920
2954 TGAAGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCGGTCAACAAT 3003
920 alValValThrLysSerAlaThrGlyAsnPheThrLeuGlnValAlaAsp 936
3004 ACCGGCAACGAACCTGCAAGCCTCGAACAATTGACGGTAGTGGAAGGAAA 3053
937 LysThrGlyGluProThrLys...AsnGluLeuThrLeuPheAspAla.. 951
3054 AGACAACAAACCGCTGTCCGAAAACCTTAATTTCACCCTGCAAAACGAAC 3103
952 SerAsnAlaThrArgAsnAsnLeuAsnValSerLeuValGlyAsnT 967
3104 ACGTCGATGCCGGCGCGTGGCGTTACCAACTCATCCGCAAAGACGGCGAG 3153
967 hrValAspLeuGlyAlaTrpLysTyrLysLeuArgAsnValAsnGlyArg 983
3154 TTCCGCCTGCATAATCCG..... 3171
984 TyrAspLeuTyrAsnProGluValGluLysArgAsnGlnThrValAspTh 1000
3171 3171
1000 rThrAsnIleThrThrProAsnAsnIleGlnAlaAspValProSerValP 1017
3171 3171
1017 roSerAsnAsnGluGluIleAlaArgValGluThrProValProProPro 1033
3171 3171
1034 AlaProAlaThrProSerGluThrThrGluThrValAlaGluAsnSerLy 1050
3171 3171
1050 sGlnGluSerLysThrValGluLysAsnGluGlnAspAlaThrGluThrT 1067
3171 3171
1067 hrAlaGlnAsnGlyGluValAlaGluGluAlaLysProSerValLysAla 1083
3171 3171
1084 AsnThrGlnThrAsnGluValAlaGlnSerGlySerGluThrGluGluTh 1100
3172 GTCAAAGAACAAGAGCTTTCCGACAAACTCGGCAAGG 3208
1100 rGlnThrThrGluIleLysGluThrAlaLysValGluLysGluGluLysA 1117
3209 CAGAAGCCAAAAAACAG..... 3225
1117 laLysValGluLysGluGluLysAlaLysValGluLysAspGluIleGln 1133
3225 3225
1134 GluAlaProGlnMetAlaSerGluThrSerProLysGlnAlaLysProAl 1150
3225 3225
1150 aProLysGluValSerThrAspThrLysValGluGluThrGlnValGlnA 1167
3225 3225
1167 laGlnProGlnThrGlnSerThrThrValAlaAlaAlaGluAlaThrSer 1183
3226 GCGGAAAAAGACAACGC 3242
1184 ProAsnSerLysProAlaGluGluThrGlnProSerGluLysThrAsnAl 1200
3243 GCAAAGCCTTGACGCGCTGATTGCG..... 3267
1200 aGluProValThrProValValSerLysAsnGlnThrGluAsnThrThrA 1217
3268 GCCGGGCGCGATGCCGTCGAAAAGACA 3294
1217 spGlnProThrGluArgGluLysThrAlaLysValGluThrGluLysThr 1233
3295 GAAAGCGTTGCCGAACCGGCCCGGCAGGCAGGC..... 3327
1234 GlnGluProProGlnValAlaSerGlnAlaSerProLysGlnGluGlnSe 1250
3328 GGGGAAAATGTC...GGCA 3343
1250 rGluThrValGlnProGlnAlaValLeuGluSerGluAsnValProThrV 1267

```
3344 TTATGCAGGCGGAGGAAGAGAAAAAACGGGTGCAGGCGGATAAAGAC... 3390
1267 alAsnAsnAlaGluGluValGlnAlaGlnLeuGlnThrGlnThrSerAla 1283

3390 ..................................................... 3390
1284 ThrValSerThrLysGlnProAlaProGluAsnSerIleAsnThrGlySe 1300

3391 ....ACCGCCTTGGCGAAACAGCGCGAAGCGGAAACCCGGCCGGCTACC. 3435
1300 rAlaThrAlaIleThrGluThrAlaGluLysSerAspLysProGlnThrG 1317

3435 ..................................................... 3435
1317 luThrAlaAlaSerThrGluAspAlaSerGlnHisLysAlaAsnThrVal 1333

3436 ............................ACCGCCTTCCCCCGCGCCCG 3455
1334 AlaAspAsnSerValAlaAsnAsnSerGluSerSerGluProLysSerAr 1350

3456 CCGCGCCCGCCGGGATTTGCCGCAACTGCAA..................... 3486
1350 gArg...ArgArgSerIleSerGlnProGlnGluThrSerAlaGluGluT 1366

3486 ..................................................... 3486
1366 hrThrAlaAlaSerThrAspGluThrThrIleAlaAspAsnSerLysArg 1382

3487 CCCCAACCGCAGCCCCAACCGCAG........................... 3510
1383 SerLysProAsnArgArgSerArgArgSerValArgSerGluProThrVa 1399

3511 ................................CGCGACCTGATCAGCC 3526
1399 lThrAsnGlySerAspArgSerThrValAlaLeuArgAspLeuThrSerT 1416

3527 GTTATGCCAATAGCGGTTTGAGTGAATTTTCCGCC................ 3561
1416 hrAsnThrAsnAlaValIleSerAspAlaMetAlaLysAlaGlnPheVal 1432

3562 ACGCTCAACAGCGTTTTCGCCGTACAGGACGAATTAGACCGCGTA...TT 3608
1433 AlaLeuAsnValGlyLysAlaValSerGlnHisIleSerGlnLeuGluMe 1449

3609 TGCCGAAGACCGCCGCAACGCCGTTTGGACAAGCGGCATCCGGGACACCA 3658
1449 tAsnAsnGluGlyGlnTyrAsnValTrpValSerAsnThrSerMetAsnG 1466

3659 AACACTACCGTTCGCAAGATTTCCGCGCCTACCGCCAACAAACCGACCTG 3708
1466 luAsnTyrSerSerSerGlnTyrArgArgPheSerSerLysSerThrGln 1482

3709 CGCCAAATCGGTATGCAGAAAAACCTCGGCAGCGGG...CGCGTCGGCAT 3755
1483 ThrGlnLeuGlyTrpAspGlnThrIleSerAsnAsnValGlnLeuGlyGl 1499

3756 CCTGTTTTCGCACAACCGGACCGAAAACACCTTCGACGACGGCATCGGCA 3805
1499 yValPheThrTyrValArgAsnSerAsnAsnPheAspLysAlaSerSerL 1516

3806 ACTCGGCACGGCTTGCCCACGGCGCCGTTTTCGGGCAATACGGCATCGAC 3855
1516 ysAsnThr...LeuAlaGlnValAsnPheTyrSerLysTyrTyrAlaAsp 1531

3856 ...AGGTTCTACATCGGCATCAGCGCGGGCGCGGGTTTTAGCAGCGGCAG 3902
1532 AsnHisTrpTyrLeuGlyIleAspLeuGlyTyrGlyLysPheGlnSerAs 1548

3903 CCTTTCAGACGGCATCGGAGGCAAAATCCGCCGCCGCGTGCTGCATTACG 3952
1548 nLeuLysThrAsnHisAsnAlaLysPheAlaArgHisThrAlaGlnPheG 1565

3953 GCATTCAGGCACGATACCGCGCCGGTTTCGGCGGATTCGGCATCGAACCG 4002
1565 lyLeuThrAlaGlyLysAlaPheAsnLeuGlyAsnPheGlyIleThrPro 1581

4003 CACATCGGCGCAACGCGCTATTTCGTCCAAAAAGCGGATTACCGCTACGA 4052
1582 IleValGlyValArgTyrSerTyrLeuSerAsnAlaAsnPheAlaLeuAl 1598

4053 AAACGTCAATATCGCCACCCCCGGCCTTGCATTCAACCGCTACCGCGCGG 4102
1598 aLysAspArgIleLysValAsnProIleSerValLysThrAlaPheAlaG 1615

4103 GCATTAAGGCAGATTATTCATTCAAACCGGCGCAACACATTTCCATCACG 4152
1615 lnValAspLeuSerTyrThrTyrHisLeuGlyGlu...PheSerValThr 1630

4153 CCTTATTTGAGCCTGTCCTAT...ACCGATGCCGCTTCGGGCAAAGTCCG 4199
1631 ProIleLeuSerAlaArgTyrAspThrAsnGlnGlySerGlyLysIle.. 1646

4200 AACACGCGTCAATACCGCCGTATTGGCTCAGGATTTCGGCAAAACCCGCA 4249
1647 ....AsnValAsnGlnTyrAspPheAlaTyrAsnValGluAsnGlnGlnG 1662

4250 GTGCGGAATGGGGCGTAAACGCCGAAATCAAAGGTTTCACGCTGTCCCTC 4299
1662 lnTyrAsnAlaGlyLeuLysLeuLysTyrHisAsnValLysLeuSerLeu 1678

4300 CACGCTGCCGCCGCCAAAGGCCCGCAACTGGAAGCGCAACACAGCGCGGG 4349
1679 IleGlyGlyLeuThrLysAlaLysGlnAlaGluLysGlnLysThrAlaGl 1695

4350 CATCAAATTAGGCTAC 4365
1695 uLeuLysLeuSerPhe 1700
```

seq_name: /cgn2_6/ptodata/1/iaa/backfiles1.pep:5268270-2

seq_documentation_block:
;Patent No. 5268270
; APPLICANT: Meyer, Thomas F.;Halter, Roman;Pohlner, Johnnnes
; TITLE OF INVENTION: PROCESS FOR PRODUCING PROTEINS USING GRAM
;NEGATIVE HOST CELLS
; NUMBER OF SEQUENCES: 6
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/171,872
; FILING DATE: 01-JUL-1987
;SEQ ID NO:2:
; LENGTH: 1507
5268270-2

alignment_scores:
Quality: 1532.50 Length: 1709
Ratio: 1.770 Gaps: 43
Percent Similarity: 50.673 Percent Identity: 27.092

alignment_block:
US-09-303-518D-649 x 5268270-2 ..

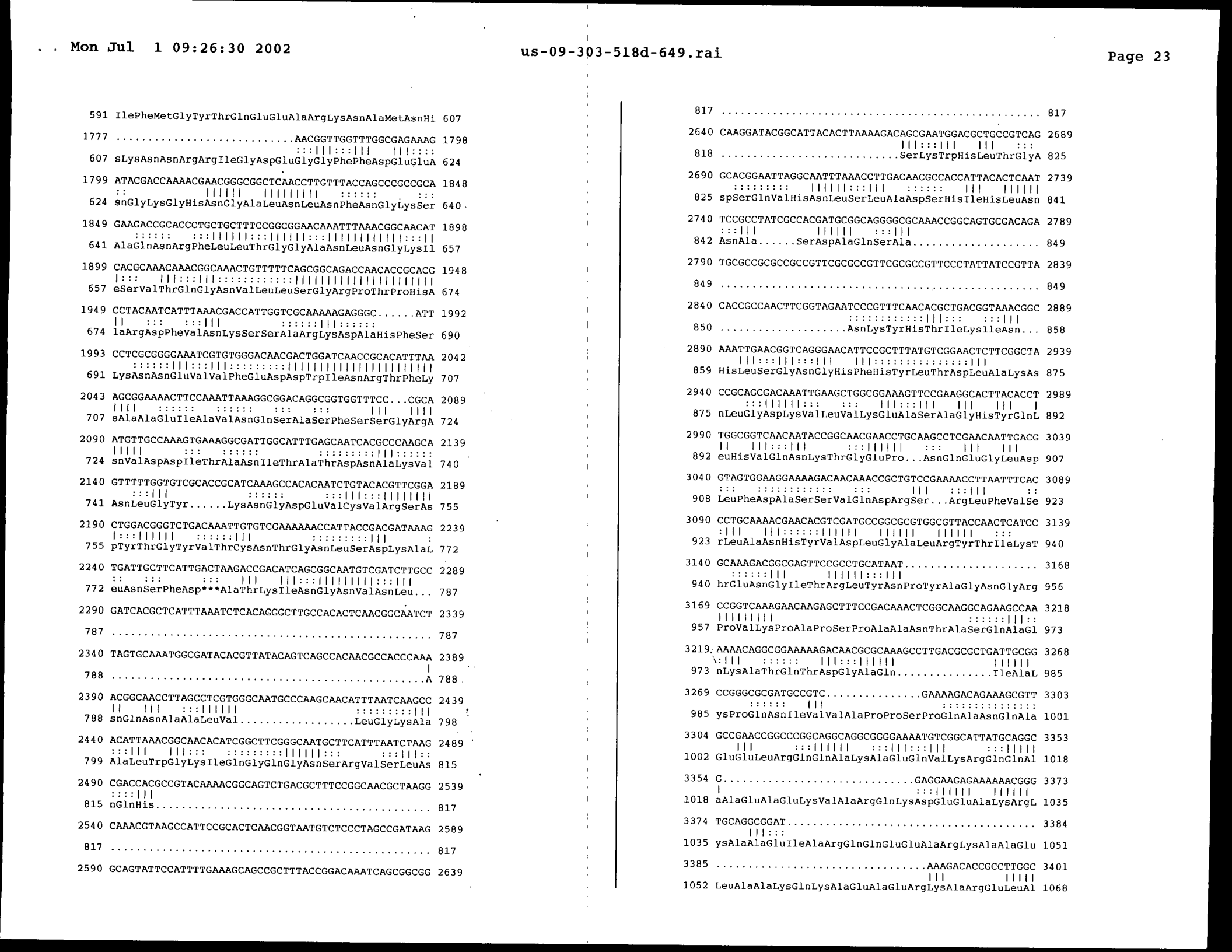
Align seg 1/1 to: 5268270-2 from: 1 to: 1507

```
  49 AAAACCGGCCGCATCCGCTTCTCGCCTGCTTACTTAGCCATATGCCTGTC 98
   2 LysAlaLysArgPheLysIleAsnAlaIleSerLeuSerIlePheLeuAl 18

  99 GTTCGGCATTCTTCCCCAAGCCTGGGCGGGACACACTTATTTCGGCATCA 148
  18 aTyrAlaLeuThrProTyrSerGluAlaAlaLeuValArgAspAspValA 35

 149 ACTACCAATACTATCGCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTC 198
  35 spTyrGlnIlePheArgAspPheAlaGluAsnLysGlyLysPhePheVal 51
```





```
3402 GAAACAGCGCGAAGCGGAAACCCGGCCGGCTACCACCGCCTTCCCCCGCG 3451
     |:::|||:::      |||::         :::::::|||   |||:::
1068 aArgGlnLysAlaGluGluAlaSerHisGlnAlaAsnAlaLysProLysA 1085

3452 CCCGCCGCGCCCGCCGGGATTTGCCGCAA..................... 3480
       ||||||   |||      ||||||:::
1085 rgArgArgArgArgAlaIleLeuProArgProProAlaProValPheSer 1101

3480 ..................................................... 3480

1102 LeuAspAspTyrAspAlaLysAspAsnSerGluSerSerIleGlyAsnLe 1118

3480 ..................................................... 3480

1118 uAlaArgValProIleArgMetGlyArgGluLeuIleAsnAspTyrGluG 1135

3480 ..................................................... 3480

1135 luIleProLeuGluGluLeuGluAspGluAlaGluGluGluArgArgGln 1151

3480 ..................................................... 3480

1152 AlaThrGlnPheHisSerLysSerArgAsnArgArgAlaIleSerSerGl 1168

3480 ..................................................... 3480

1168 uProSerSerAspGluAspAlaSerGluSerValSerThrSerAspLysH 1185

3480 ..................................................... 3480

1185 isProGlnAspAsnThrGluLeuHisGluLysValGluThrAlaGlyLeu 1201

3481 ............CTGCAACCCCAACCGCAGCCCCAACCGCAGCGCGACCT 3518
                 ||||||::   |||        |||   |||
1202 GlnProArgAlaAlaGlnProArgThrGlnAlaAlaAlaGlnAlaAspAl 1218

3519 GATCAGCCGTTATGCCAATAGCGGTTTGAGTGAATTTTCCGCCACGCTCA 3568
      :::|||      :::|||||||:::||||||||:::      |||:::   :
1218 aValSerThrAsnThrAsnSerAlaLeuSerAspAlaMetAlaSerThrG 1235

3569 ACAGCGTT.................TTCGCCGTACAGGACGAATTAGAC 3600
     ::|||:::                 :::   ::::::         :::
1235 lnSerIleLeuLeuAspThrGlyAlaTyrLeuThrArgHisIleAlaGln 1251

3601 CGCGTATTTGCCGAAGACCGCCGCAACGCCGTTTGGACAAGCGGCATCCG 3650
     :::      |||:::   ::::::|||:::||||||   ::::::
1252 LysSerArgAlaAspAlaGluLysAsnSerValTrpMetGluAsnThrGl 1268

3651 GGACACCAAACACTACCGTTCGCAAGATTTCCGCGCCTACCGCCAACAAA 3700
           :::   |||   |||   ::::::|||   :::   ::::::
1268 yTyrGlyArgAspTyrAlaSerAlaGlnTyrArgArgPheSerSerLysA 1285

3701 CCGACCTGCGCCAAATCGGTATGCAGAAAAACCTCGGCAGCGGG...CGC 3747
               |||||||||:::::::::::|||:::::::::   :::
1285 rgThrGlnThrGlnIleGlyIleAspArgSerLeuSerGluAsnMetGln 1301

3748 GTCGGCATCCTGTTTTCGCACAACCGGACCGAAAACACCTTCGACGACGG 3797
     :::|||   ::::::::::::::   :::::::::|||||||||:::::
1302 IleGlyGlyValLeuThrTyrSerAspSerGlnHisThrPheAspGlnAl 1318

3798 CATCGGCAACTCGGCACGGCTTGCCCACGGCGCCGTTTTCGGGCAATACG 3847
     :   |||::::::      ::::::::::::   :::::::|||:::|||
1318 a...GlyGlyLysAsnThrPheValGlnAlaAsnLeuTyrGlyLysTyrT 1334

3848 GCATC...GACAGGTTCTACATCGGCATCAGCGCGGGCGCGGGTTTTAGC 3894
       :::   |||   :::|||:::::   :::   |||||||||
1334 yrLeuAsnAspAlaTrpTyrValAlaGlyAspIleGlyAlaGlySerLeu 1350

3895 AGCGGCAGCCTTTCAGACGGCATCGGAGGCAAAATCCGCCGCCGCCGTGCT 3944
        :::   |||:::          ::::::::::::|||       ::
1351 ArgSerArgLeuGlnThrGlnGlnLysAlaAsnPheAsnArgThrSerIl 1367

3945 GCATTACGGCATTCAGGCACGATACCGCGCCGGTTTCGGCGGATTCGGCA 3994
     ::::  |||:::                    ::::::   |||   |
1367 eGlnThrGlyLeuThrLeuGlyAsnThrLeuLysIleAsnGlnPheGluI 1384

3995 TCGAACCGCACATCGGCGCAACGCGCTATTTCGTCCAAAAAGCGGATTAC 4044
     ||   |||      |||              :::::::::|||||||||
1384 leValProSerAlaGlyIleArgTyrSerArgLeuSerSerAlaAspTyr 1400

4045 CGCTACGAAAACGTCAATATCGCCACCCCCGGCCTTGCATTCAACCGCTA 4094
     :::      :::   ::::::   :::   :::::::|||   :::
1401 LysLeuGlyAspAspSerValLysValSerSerMetAlaValLysThrLe 1417

4095 CCGCGCGGGCATTAAGGCAGATTATTCATTCAAACCGGCGCAACACATTT 4144
         ||||||:::          |||   ||||||   :::   :::::::
1417 uThrAlaGlyLeuAspPheAlaTyrArgPheLys...ValGlyAsnLeuT 1433

4145 CCATCACGCCTTATTTGAGCCTGTCCTATACCGATGCCGCTTCGGGCAAA 4194
     :::::   |||   ||||||   :::|||       |||       ||||||
1433 hrValLysProLeuLeuSerAlaAlaTyrPhe...AlaAsnTyrGlyLys 1448

4195 GTCCGAACACGCGTCAATACCGCCGTATTGGCTCAGGATTTCGGCAAAAC 4244
           ::::::|||:::         :::|||              :::
1449 GlyGlyValAsnValGlyGlyLysSerPheAlaTyrLysAlaAspAsnGl 1465

4245 CCGCAGTGCGGAATGGGGCGTAAACGCCGAAATCAAAGGTTTCACGCTGT 4294
       ::::::   :::   ||||||            ::::::   ||||||:
1465 nGlnGlnTyrSerAlaGlyValAlaLeuLeuTyrArgAsnValThrLeuA 1482

4295 CCCTCCACGCTGCCGCCGCCAAAGGCCCGCAACTGGAAGCGCAACACAGC 4344
     ::::::::::::::   :::||||||   |||||||||   |||   |||
1482 snValAsnGlySerIleThrLysGlyLysGlnLeuGluLysGlnLysSer 1498

4345 GCGGGCATCAAATTAGGCTACCGCTGG 4371
     :::   |||||||:::      |||:::
1499 GlyGlnIleLysIleGlnIleArgPhe 1507
```

```
seq_name: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-296-791-6

seq_documentation_block:
; Sequence 6, Application US/08296791
; Patent No. 6245337
;  GENERAL INFORMATION:
;    APPLICANT:  St. Geme III, Joseph W.
;    APPLICANT:  Falkow, Stanley
;    TITLE OF INVENTION:  Haemophilus Adherence and Penetration
;    TITLE OF INVENTION:  Protein
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Flehr, Hohbach, Test, Albritton & Herbert
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/296,791
;      FILING DATE:  25-AUG-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Trecartin, Richard F.
;      REGISTRATION NUMBER:  31,801
;      REFERENCE/DOCKET NUMBER:  A-59941/RFT/RMS
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
```

```
;       LENGTH:  1848 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  unknown
US-08-296-791-6


alignment_scores:
            Quality: 1486.00              Length:   1916
              Ratio:  1.490                Gaps:     58
 Percent Similarity:  52.035   Percent Identity: 24.843

alignment_block:
 US-09-303-518D-649 x US-08-296-791-6   ..

 Align seg 1/1  to: US-08-296-791-6  from: 1  to: 1848

      64 CGCTTCTCGCCTGCTTACTTAGCCATATGCCTGTCGTTCGGCATTCTTCC 113
         :::|||:::      ::::::|||:::   ::::::::::::::   ||
       5 LysPheLysLeuAsnPheIleAlaLeuThrValAlaTyrAlaLeuThrPr 21

     114 CCAAGCCTGGGCGGGACACACTTATTTCGGCATCAACTACCAATACTATC 163
         |   :::   |||:::   :::         ::::::||||||   :::|
      21 oTyrThrGluAlaAlaLeuValArgAspAspValAspTyrGlnIlePheA 38

     164 GCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATT 213
         |||||||||||||||||||||||||||||:::|||||||||   ::::::
      38 rgAspPheAlaGluAsnLysGlyLysPheSerValGlyAlaThrAsnVal 54

     214 GAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAATCAATGACAAAAGC 263
         ||||||   :::||||||::::::   :::|||:::::::::   :::::
      55 GluValArgAspLysLysAsnGlnSerLeuGlySerAlaLeuProAsnGl 71

     264 C...CCGATGATTGATTTTTCTGTGGTGTCG...CGTAACGGCGTGGCGG 307
         :   ||||||||||||||||||||||||:::      :::   :::|||:
      71 yIleProMetIleAspPheSerValValAspValAspLysArgIleAlaT 88

     308 CATTGGTGGGCGATCAATATATTGTGAGCGTGGCACAT...AACGGCGGC 354
         ::||||||:::   ||||||:::|||:::|||   |||   ::::::|||
      88 hrLeuValAsnProGlnTyrValValGlyValLysHisValSerAsnGly 104

     355 TATAACAACGTTGATTTTGGT..............GCGGAAGGAAGAAA 389
         :::::::::   ||||||                  ::::::   ||
     105 ValSerGluLeuHisPheGlyAsnLeuAsnGlyAsnMetAsnAsnGlyAs 121

     390 TCCCGATCAACATCGTTTTACT.................TATAAAATTG 421
         |      :::||||||   :::                 |||      |
     121 nAlaLysSerHisArgAspValSerSerGluGluAsnArgTyrTyrThrV 138

     422 TGAAACGGAATAATTAT.........................AAAGCA 444
         ||:::::::||||||:::                        |||
     138 alGluLysAsnAsnPheProThrGluAsnValThrSerPheThrLysGlu 154

     445 GGGACTAAAGGCCATCCTTATGGCGGCGATTATCATATGCCGCGTTTGCA 494
                 ::::::          ||||||:::|||||||||||||
     155 GluGlnAspAlaGlnLysArgArgGluAspTyrTyrMetProArgLeuAs 171

     495 TAAATTTGTCACAGATGCAGAACCTGTTGAAATGACCAGTTATATGGATG 544
         ||||||||||||:::::   |||:::|||   ::::::      ::::
     171 pLysPheValThrGluValAlaProIleGluAlaSerThrAlaAsnAsnA 188

     545 GGCGG...AAATATATCGATCAAAATAATTACCCTGACCGTGTTCGTATT 591
         :::::   :::|||   ::::::::::::|||||      ||||||:::
     188 snLysGlyGluTyrAsnAsnSerAspLysTyrProAlaPheValArgLeu 204

     592 GGGGCAGGCAGGCAATAT............................TG 611
         |||:::|||   |||:::
     205 GlySerGlyThrGlnPheIleTyrLysLysGlySerArgTyrGlnLeuIl 221

     612 GCGATCTGATGAAGATGAGCCCAATAACCGCGAAAGTTCATATCAT.... 657
             ::::::::::|||:::   :::|||         ::::::
     221 eLeuThrGluLysAspLysGlnGlyAsnLeuLeuArgAsnTrpAspValG 238

     658 ...................ATTGCAAGTGCGTATTCTTGGCTCGTTGGT 687
                                    :::::::::||||||::::::   ::::::
     238 lyGlyAspAsnLeuGluLeuValGlyAsnAlaTyrThrTyrGlyIleAla 254

     688 GGCAATACCTTTGCACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGG 737
         |||:::   :::      |||   ::::::::::|||   ::::::::::||
     255 GlyThrProTyrLysValAsnHisGluAsnAsnGlyLeuIleGlyPheGl 271

     738 TAGTGAAAAAAATTAAACATAGC...CCATATGGTTTTTTACCAACAGGA. 783
         |:::::::|||   :::||||||   |||   |||:::|||
     271 yAsnSerLysGluGluHisSerAspProLysGlyIleLeuSerGlnAspP 288

     784 .............GGCTCATTTGGCGACAGTGGCTCACCAATGTTTATC 819
                     :::   :::||||||||||||||||||||:::|||:::
     288 roLeuThrAsnTyrAlaValLeuGlyAspSerGlySerProLeuPheVal 304

     820 TATGATGCCCAAAAGCAAAAGTGGTTAATTAATGGGGTATTGCAAACGGG 869
         ||||||   :::|||   ||||||||||:::   |||       :::
     305 TyrAspArgGluLysGlyLysTrpLeuPheLeuGlySerTyrAspPheTr 321

     870 CAACCCCTATATAGGAAAAAGCAATGGCTTCCAGCTGGTTCGTAAAGATT 919
               |||       ||||||                 :::   |||
     321 pAlaGlyTyrAsnLysLysSerTrpGlnGluTrpAsnIleTyrLysHisG 338

     920 GGTTCTATGATGAAATCTTTGCTGGAGATACCCATTCAGTATTCTACGAA 969
           |||   ::::::|||:::       :::   :::|||:::
     338 luPheAlaGluLysIleTyr......GlnGlnTyrSerAlaGlySerLeu 352

     970 CCACGTCAAAATGGGAAATACTCTTTTAACGACGATAATAATGGCACAGG 1019
               :::|||   :::|||::::::            :::::||||
     353 IleGlySerAsnThrGlnTyrThrTrp...........GlnAlaThrGl 365

    1020 AAAAATCAATGCCAAACATGAACACAATTCTCTGCCTAATAGATTAAAAA 1069
         |                                              :::|
     365 y.........................................SerT 367

    1070 CACGAACCGTT............CAATTGTTTAATGTTTCTTTATCCGAG 1107
         ||   |||:::            :::   ::::::|||:::|||::::::
     367 hrSerThrIleThrGlyGlyGluProLeuSerValAspLeuThrAsp 383

    1108 ACAGCAAGAGAACCTGTTTATCATGCTGCAGGTGGTGTCAACAGTTATCG 1157
                 :::|||
     384 GlyLysAspLysPro.................................. 388

    1158 ACCCAGACTGAATAATGGAGAAAATATTTCCTTTATTGACGAAGGAAAAG 1207
               |||:::|||::::::|||::::::      :::|||:::|
     389 ..........AsnHisGlyLysSerIleThrLeu......LysGlySerG 400

    1208 GCGAATTGATACTTACCAGCAACATCAATCAAGGTGCTGGAGGATTATAT 1257
         ||   |||   |||:::::::::|||:::||||||||||||||||||:::
     400 lyThrLeuThrLeuAsnAsnHisIleAspGlnGlyAlaGlyGlyLeuPhe 416

    1258 TTCCAAGGAGATTTTACGGTC...TCGCCTGAAAATAACGAAACTTGGCA 1304
         |||:::||||||:::   |||   :::   :::::::::   |||||||::
     417 PheGluGlyAspTyrGluValLysGlyThrSerAspSerThrThrTrpLy 433

    1305 AGGCGCGGGCGTTCATATCAGTGAAGACAGTACCGTTACTTGGAAAGTAA 1354
         :|||||||||||||   :::::::::   :::|||||||||||||||||:
     433 sGlyAlaGlyValSerValAlaAspGlyLysThrValThrTrpLysValH 450

    1355 ACGGCGTGGCAAACGACCGCCTGTCCAAAATCGGCAAAGGCACGCTGCAC 1404
         :::::          ||||||||||:::|||||||||||||||||||
     450 isAsnProLysTyrAspArgLeuAlaLysIleGlyLysGlyThrLeuVal 466

    1405 GTTCAAGCCAAAGGGGAAAACCAAGGCTCGATCAGCGTGGGCGACGGTAC 1454
         |||:::::::|||||:::|||:::|||   ::::::|||||||||||||||
     467 ValGluGlyLysGlyLysAsnGluGlyLeuLeuLysValGlyAspGlyTh 483

    1455 AGTCATTTTGGATCAGCAGGCAGACGATAAAGGCAAAAAACAAGCCTTTA 1504
         ||||||||||   |||:::||||||   ::::::|||   ||||||||||
     483 rValIleLeuLysGlnLysAlaAspAlaAsnAsnLysValGlnAlaPheS 500

    1505 GTGAAATCGGCTTGGTCAGCGGCAGGGGTACGGTGCAACTGAATGCCGAT 1554
         ||:::::|||:::||||||||||||:::|||:::   ||||||   |||
```


500 erGlnValGlyIleValSerGlyArgSerThrLeuValLeuAsnAspAsp 516
1555 AATCAGTTCAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTT 1604
517 LysGlnValAspProAsnSerIleTyrPheGlyPheArgGlyGlyArgLe 533
1605 GGATTTAAACGGGCATTCGCTTTCGTTCCACCGTATTCAAAATACCGATG 1654
533 uAspLeuAsnGlyAsnSerLeuThrPheAspHisIleArgAsnIleAspA 550
1655 AAGGGGCGATGATTGTCAACCACAATCAAGACAAAGAATCCACCGTTACC 1704
550 spGlyAlaArgValValAsnHisAsnMetThrAsnThrSerAsnIleThr 566
1705 ATTACAGGC..... 1713
567 IleThrGlyGluSerLeuIleThrAsnProAsnThrIleThrSerTyrAs 583
1713 1713
583 nIleGluAlaGlnAspAspAspHisProLeuArgIleArgSerIleProT 600
1713 1713
600 yrArgGlnLeuTyrPheAsnGlnAspAsnArgSerTyrTyrThrLeuLys 616
1713 1713
617 LysGlyAlaSerThrArgSerGluLeuProGlnAsnSerGlyGluSerAs 633
1714 AATAAAGATATTGCTACAACCGGCA 1738
633 nGluAsnTrpLeuTyrMetGlyArgThrSerAspAlaAlaLysArgAsnV 650
1739 ATAACAACAGCTTGGATAGCAAAAAAGAAATTGCCTACAACGGTTGGTTT 1788
650 alMetAsnHisIleAsnAsnGluArgMetAsnGlyPheAsnGlyTyrPhe 666
1789 GGCGAGAAAGATACGACCAAAACG...AACGGGCGGCTCAACCTTGTTTA 1835
667 GlyGluGluGluThrLysAlaThrGlnAsnGlyLysLeuAsnValThrPh 683
1836 CCAGCCCGCCGCAGAAGACCGCACCCTGCTGCTTTCCGGCGGAACAAATT 1885
683 eAsnGlyLysSerAspGlnAsnArgPheLeuLeuThrGlyGlyThrAsnL 700
1886 TAAACGGCAACATCACGCAAACAAACGGCAAACTGTTTTTCAGCGGCAGA 1935
700 euAsnGlyAspLeuAsnValGluLysGlyThrLeuPheLeuSerGlyArg 716
1936 CCAACACCGCACGCCTACAATCATTTAAACGACCATTGGTCGCAAAAAGA 1985
717 ProThrProHisAlaArgAspIleAlaGlyIleSerSerThrLysLysAs 733
1986 GGGC......ATTCCTCGCGGGGAAATCGTGTGGGACAACGACTGGATCA 2029
733 pProHisPheThrGluAsnAsnGluValValValGluAspAspTrpIleA 750
2030 ACCGCACATTTAAAGCGGAAAACTTCCAAATTAAAGGCGGACAGGCGGTG 2079
750 snArgAsnPheLysAlaThrThrMetAsnValThrGlyAsnAlaSerLeu 766
2080 GTTTCC...CGCAATGTTGCCAAAGTGAAAGGCGATTGGCATTTGAGCAA 2126
767 TyrSerGlyArgAsnValAlaAsnIleThrSerAsnIleThrAlaSerAs 783
2127 TCACGCCCAAGCAGTTTTTGGTGTCGCACCGCATCAAAGCCACACAATCT 2176
783 nAsnAlaGlnValHisIleGlyTyr......LysThrGlyAspThrValC 798
2177 GTACACGTTCGGACTGGACGGGTCTGACAAATTGTGTCGAAAAAACCATT 2226
798 ysValArgSerAspTyrThrGlyTyrValThrCysHisAsnSerAsnLeu 814
2227 ACCGACGATAAAGTGATTGCTTCATTGACTAAGACCGACATCAGCGGCAA 2276
815 Ser...GluLysAlaLeuAsnSerPheAsnProThrAsnLeuArgGlyAs 830
2277 TGTCGATCTTGCCGATCACGCTCATTTAAATCTCACAGGGCTTGCCACAC 2326
830 nValAsnLeuThrGluAsnAlaSerPheThrLeu...GlyLysAlaAsnL 846
2327 TCAACGGCAATCTTAGTGCAAATGGCGATACACGTTATACAGTCAGCCAC 2376
846 euPheGlyThrIleGlnSerIleGlyThrSerGlnValAsnLeuLysGlu 862
2377 AACGCC.....ACCCAAAACGGCAACCTTAGCCTCGTGGGC.. 2412
863 AsnSerHisTrpHisLeuThrGlyAsnSerAsnValAsnGlnLeuAsnLe 879
2413 AATGCCCAA.....G 2422
879 uThrAsnGlyHisIleHisLeuAsnAlaGlnAsnAspAlaAsnLysValT 896
2423 CAACATTTAATCAAGCCACATTAAACGGCAACACATCGGCTTCGGGCAAT 2472
896 hrThrTyrAsnThrLeuThrValAsn.....SerLeuSerGlyAsn 909
2473 GCTTCATTT.....AATCTAAGCGACCACGC 2498
910 GlySerPheTyrTyrTrpValAspPheThrAsnAsnLysSerAsnLysVa 926
2499 CGTACAAAAC.....GGCAGTCTGACGCTT..... 2523
926 lValValAsnLysSerAlaThrGlyAsnPheThrLeuGlnValAlaAspL 943
2524 TCCGGC 2529
943 ysThrGlyGluProAsnHisAsnGluLeuThrLeuPheAspAlaSerAsn 959
2530 AACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACGGTAATGTCTCCCT 2579
960 AlaThrArgAsnAsnLeuGluValThrLeuAlaAsnGlySerVal..... 974
2580 AGCCGATAAGGCAGTATTCCATTTTGAAAGCAGCCGCTTTACCGGACAAA 2629
975 AspArgGlyAlaTrpLysTyrLysLeuArgAsnValAsnGlyArgT 990
2630 TCAGCGGCGGCAAGGATACGGCATTACACTTAAAAGACAGCGAATGGACG 2679
990 yrAsp.....LeuTyrAsnProGlu..... 996
2680 CTGCCGTCAGGCACGGAATTAGGCAATTTAAACCTTGACAACGCCACCAT 2729
997 ValGluLysArgAsnGlnThrValAspThrThrAsnIl 1009
2730 TACACTCAATTCCGCCTATCGCCACGATGCGGCAGGGGCGCAAACCGGCA 2779
1009 eThrThrProAsnAspIleGlnAlaAspAlaProSerAlaGlnSerAsnA 1026
2780 GT..... 2781
1026 snGluGluIleAlaArgValGluThrProValProProProAlaProAla 1042
2782 GCGACAGATGCGCCGCGCCGCCGTTCGCGCCGTTC 2816
1043 ThrGluSerAlaIleAlaSerGluGlnProGluThrArgProAlaGluTh 1059
2817 GCGCCGTTCCCTATTATCCGTTACA..... 2841
1059 rAlaGlnProAlaMetGluGluThrAsnThrAlaAsnSerThrGluThrA 1076
2841 2841
1076 laProLysSerAspThrAlaThrGlnThrGluAsnProAsnSerGluSer 1092

```
2842 ..................................CCGCCAACTTCGGT 2855
                                       ||||||   :::
1093 ValProSerGluThrThrGluLysValAlaGluAsnProProGlnGluAs 1109

2856 AGAATCC.....................CGTTTCAACACGCTGACGGTAA 2884
      |||:::                          :::          |
1109 nGluThrValAlaLysAsnGluGlnGluAlaThrGluProThrProGlnA 1126

2885 ACGGCAAATTG.............................AACGGTCAG 2904
     |||||::::::                           |||   |||
1126 snGlyGluValAlaLysGluAspGlnProThrValGluAlaAsnThrGln 1142

2905 GGAACATTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGACAAATT 2954
        :::         :::||||||           ::::::::::
1143 ThrAsnGluAlaThrGlnSerGlu.........GlyLysThrGluGluTh 1156

2955 GAAGCTGGCGGAAAGT...TCCGAAGGCACTTACACCTTGGCGGTCAACA 3001
      :::   ||||||:::   ||||||   |||   :::::::::|||::::
1156 rGlnThrAlaGluThrLysSerGluProThrGluSerValThrValSerG 1173

3002 ATACCGGCAACGAACCT..............GCAAGCCTCGAACAATTG 3036
     ::       |||:::|||                |||   |||:::
1173 lu......AsnGlnProGluLysThrValSerGlnSerThrGluAspLys 1187

3037 ACGGTAGTGGAAGGAAAAGAC.......................... 3057
     :::||||||||||   ::::::
1188 ValValValGluLysGluGluLysAlaLysValGluThrGluGluThrGl 1204

3057 ..................................................... 3057

1204 nLysAlaProGlnValThrSerLysGluProProLysGlnAlaGluProA 1221

3057 ..................................................... 3057

1221 laProGluGluValProThrAspThrAsnAlaGluGluAlaGlnAlaLeu 1237

3057 ..................................................... 3057

1238 GlnGlnThrGlnProThrThrValAlaAlaAlaGluThrThrSerProAs 1254

3058 .AACAAACCGCTGTCCGAAAACCTTAATTTCACCCTGCAAAACGAAC... 3103
      :::|||||||   :::||| |||                |||||||||
1254 nSerLysProAlaGluGlu.ThrGlnGlnProSerGluLysThrAsnAla 1270

3104 .........ACGTCGATGCCGGCGCGTGGCGTTACCAACTCATCCGCAAA 3144
              |||   :::   ::::::::::::::::::::::   :::::
1271 GluProValThrProValValSerGluAsnThrAlaThrGlnProThrGl 1287

3145 GACGGCGAGTTCCGCCTGCATAATCCGGTCAAAGAACAAGAGCTTTCCGA 3194
     :|||   :::::::|||   :::   :::::::|||:::::::   |||
1287 uThrGluGluThrAlaLysValGluLysGluLysThrGlnGluValProG 1304

3195 CAAACTCGGCAAGGCAGAAGCCAAAAAACAGGCGGAAAAAGACAACGCGC 3244
          :::::::::::::::|||||:::::::::      :::.
1304 lnValAlaSerGlnGluSerProLysGlnGluGlnProAlaAlaLysPro 1320

3245 AAAGCCTTGACGCGCTGATTGCGGCCGGGCGCGATGCCGTCGAAAAG... 3291
     :::|||   |||          :::|||   |||   |||::::::
1321 GlnAlaGlnThr.........LysProGlnAlaGluProAlaArgGluAs 1334

3292 ..........ACAGAAAGCGTTGCCGAACCGGCCCGGCAGGCA....... 3324
              |||::::::|||:::||||||      ||||||
1334 nValLeuThrThrLysAsnValGlyGluProGlnProGlnAlaGlnProG 1351

3324 ..................................................... 3324

1351 lnThrGlnSerThrAlaValProThrThrGlyGluThrAlaAlaAsnSer 1367

3325 .....................................................GG 3326
                                                          ::
1368 LysProAlaAlaLysProGlnAlaGlnAlaLysProGlnThrGluProAl 1384

3327 CGGGGAAAATGTCGGCATTATGCAGGCGGAGGAA.............. 3360
```

```
     :   |||||||||:::   ::::::::::::|||
1384 aArgGluAsnValSerThrValAsnThrLysGluProGlnSerGlnThrS 1401

3360 ..................................................... 3360

1401 erAlaThrValSerThrGluGlnProAlaLysGluThrSerSerAsnVal 1417

3360 ..................................................... 3360

1418 GluGlnProAlaProGluAsnSerIleAsnThrGlySerAlaThrThrMe 1434

3360 ..................................................... 3360

1434 tThrGluThrAlaGluLysSerAspLysProGlnMetGluThrValThrG 1451

3360 ..................................................... 3360

1451 luAsnAspArgGlnProGluAlaAsnThrValAlaAspAsnSerValAla 1467

3360 ..................................................... 3360

1468 AsnAsnSerGluSerSerGluSerLysSerArgArgArgArgSerValSe 1484

3361 .GAGAAAAAACGGGTGCAGGCGGATAAAGACACCGCCTTGGCGAAACAGC 3409
      :::   |||:::::::::|||::::::   |||:::   :::   |||:
1484 rGlnProLysGluThrSerAlaGluGluThrThrValAlaSerThrGlnG 1501

3410 GCGAAGCGGAAACCCGGCCGGCTACCACCGCCTTCCCCCGCGCCCGCCGC 3459
     ::   :::      :::   :::::::|||      ||||||:::||||||
1501 luThrThrValAspAsnSerValSerThrProLysProArgSerArgArg 1517

3460 GCCCGCCGGGATTTGCCGCAACTGCAACCCCAACCGCAGCCCCAACCGCA 3509
     :::||||||::::::          :::   :::|||         |||
1518 ThrArgArgSerValGlnThrAsnSerTyrGluProValGluLeuProTh 1534

3510 G..................................................... 3510

1534 rGluAsnAlaGluAsnAlaGluAsnValGlnSerGlyAsnAsnValAlaA 1551

3511 ................CGCGACCTGATCAGCCGTTATGCCAATAGCGGT 3543
                     |||:::|||   |||:::   :::|||:::
1551 snSerGlnProAlaLeuArgAsnLeuThrSerLysAsnThrAsnAlaVal 1567

3544 TTGAGTGAATTTTCCGCC..............ACGCTCAACAGCGTTTT 3578
     :::|||:::      |||               :::||||||
1568 IleSerAsnAlaMetAlaLysAlaGlnPheValAlaLeuAsnValGlyLy 1584

3579 CGCCGTACAGGACGAATTAGACCGCGTA...TTTGCCGAAGACCGCCGCA 3625
      ||||||:::::::::::::::::::::   :::   ::::::   :::
1584 sAlaValSerGlnHisIleSerGlnLeuGluMetAsnAsnGluGlyGlnT 1601

3626 ACGCCGTTTGGACAAGCGGCATCCGGGACACCAAACACTACCGTTCGCAA 3675
          ||||||   |||:::          :::|||:::|||   |||:::
1601 yrAsnValTrpIleSerAsnThrSerMetAsnLysAsnTyrSerSerGlu 1617

3676 GATTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCAAATCGGTATGCA 3725
     :::::::|||   :::   :::::::::          |||:::|||   ::
1618 GlnTyrArgArgPheSerSerLysSerThrGlnThrGlnLeuGlyTrpAs 1634

3726 GAAAAACCTCGGCAGCGGG...CGCGTCGGCATCCTGTTTTCGCACAACC 3772
     ::::::::::::::::::   ::::::|||   :::|||::::::   |
1634 pGlnThrIleSerAsnAsnValGlnLeuGlyGlyValPheThrTyrValA 1651

3773 GGACCGAAAACACCTTCGACGACGGCATCGGCAACTCGGCACGGCTTGCC 3822
     ||::::::|||:::||||||   :::   ::::::::::::   ||||||
1651 rgAsnSerAsnAsnPheAspLysAlaSerSerLysAsnThr...LeuAla 1666

3823 CACGGCGCCGTTTTCGGGCAATACGGCATCGAC...AGGTTCTACATCGG 3869
     :::          :::::::::|||      |||   ::::::|||:::||
1667 GlnValAsnPheTyrSerLysTyrTyrAlaAspAsnHisTrpTyrLeuGl 1683

3870 CATCAGCGCGGGCGCGGGTTTTAGCAGCGGCAGCCTTTCAGACGGCATCG 3919
     ||||:::   |||   |||      :::::::::|||:::   :::   :
```

```
    1683 yIleAspLeuGlyTyrGlyLysPheGlnSerAsnLeuGlnThrAsnAsnA 1700

    3920 GAGGCAAAATCCGCCGCCGCGTGCTGCATTACGGCATTCAGGCACGATAC 3969
         :::::|||:::   |||::::::   :::   |||:::   |||
    1700 snAlaLysPheAlaArgHisThrAlaGlnIleGlyLeuThrAlaGlyLys 1716

    3970 CGCGCCGGTTTCGGCGGATTCGGCATCGAACCGCACATCGGCGCAACGCG 4019
               ::::::|||:::|||:::::::::|||   :::|||:::
    1717 AlaPheAsnLeuGlyAsnPheAlaValLysProThrValGlyValArgTy 1733

    4020 CTATTTCGTCCAAAAAGCGGATTACCGCTACGAAAACGTCAATATCGCCA 4069
            :::::::::::::||||||:::          :::   :::|||   :
    1733 rSerTyrLeuSerAsnAlaAspPheAlaLeuAlaGlnAspArgIleLysV 1750

    4070 CCCCCGGCCTTGCATTCAACCGCTACCGCGCGGGCATTAAGGCAGATTAT 4119
         ::      ::::::   :::         |||   :::      :::|||
    1750 alAsnProIleSerValLysThrAlaPheAlaGlnValAspLeuSerTyr 1766

    4120 TCATTCAAACCGGCGCAACACATTTCCATCACGCCTTATTTGAGCCTGTC 4169
         ::::::      ::::::   :::||||||||||||   ||||||
    1767 ThrTyrHisLeuGlyGlu...PheSerIleThrProIleLeuSerAlaAr 1782

    4170 CTATACCGATGCCGCTTCGGGCAAAGTCCGAACACGCGTCAATACCGCCG 4219
          |||   ||||||   :::|||:::   :::   :::|||::::::
    1782 gTyr...AspAlaAsnGlnGlyAsnGlyLysIleAsnValSerValTyrA 1798

    4220 TATTGGCTCAGGATTTCGGCAAAACCCGCAGTGCGGAATGGGGCGTAAAC 4269
           :::|||   :::      :::   ::::::   :::   |||::::::
    1798 spPheAlaTyrAsnValGluAsnGlnGlnGlnTyrAsnAlaGlyLeuLys 1814

    4270 GCCGAAATCAAAGGTTTCACGCTGTCCCTCCACGCTGCCGCCGCCAAAGG 4319
            :::      :::      |||||||||   ::::::   :::||||::
    1815 LeuLysTyrHisAsnValLysLeuSerLeuIleGlyGlyLeuThrLysAl 1831

    4320 CCCGCAACTGGAAGCGCAACACAGCGCGGGCATCAAATTAGGCTAC 4365
         :   |||   |||   |||   :::|||   :::||||||::::::
    1831 aLysGlnAlaGluLysGlnLysThrAlaGluValLysLeuSerPhe 1846

seq_name: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:PCT-US95-10661A-6

seq_documentation_block:
; Sequence 6, Application PC/TUS9510661A
;  GENERAL INFORMATION:
;    APPLICANT:  Washington University, et al.
;    TITLE OF INVENTION:  Haemophilus Adherence and Penetration Protein
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Flehr, Hohbach, Test, Albritton & Herbert
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/10661A
;      FILING DATE:  16-AUG-1995
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/296,791
;      FILING DATE:  25-AUG-1994
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Trecartin, Richard F.
;      REGISTRATION NUMBER:  31,801
;      REFERENCE/DOCKET NUMBER:  FP-59941/RFT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1848 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
PCT-US95-10661A-6


alignment_scores:
            Quality: 1486.00              Length:   1916
              Ratio:  1.490                Gaps:     58
 Percent Similarity:  52.035   Percent Identity: 24.843

alignment_block:
 US-09-303-518D-649 x PCT-US95-10661A-6   ..

 Align seg 1/1  to: PCT-US95-10661A-6  from: 1  to: 1848

      64 CGCTTCTCGCCTGCTTACTTAGCCATATGCCTGTCGTTCGGCATTCTTCC 113
         :::|||:::      ::::::|||:::   ::::::::::::::   ||
       5 LysPheLysLeuAsnPheIleAlaLeuThrValAlaTyrAlaLeuThrPr 21

     114 CCAAGCCTGGGCGGGACACACTTATTTCGGCATCAACTACCAATACTATC 163
         |   :::   |||:::   :::          ::::::||||||   :::|
      21 oTyrThrGluAlaAlaLeuValArgAspAspValAspTyrGlnIlePheA 38

     164 GCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATT 213
         ||||||||||||||||||||||||||||:::|||||||||   ::::::
      38 rgAspPheAlaGluAsnLysGlyLysPheSerValGlyAlaThrAsnVal 54

     214 GAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAATCAATGACAAAAGC 263
         ||||||   :::||||||::::::   :::|||:::::::::   :::::
      55 GluValArgAspLysLysAsnGlnSerLeuGlySerAlaLeuProAsnGl 71

     264 C...CCGATGATTGATTTTTCTGTGGTGTCG...CGTAACGGCGTGGCGG 307
         :   ||||||||||||||||||||||||::      :::   :::|||:
      71 yIleProMetIleAspPheSerValValAspValAspLysArgIleAlaT 88

     308 CATTGGTGGGCGATCAATATATTGTGAGCGTGGCACAT...AACGGCGGC 354
         ::||||||:::   ||||||:::|||:::|||   |||   :::::::|||
      88 hrLeuValAsnProGlnTyrValValGlyValLysHisValSerAsnGly 104

     355 TATAACAACGTTGATTTTGGT.............GCGGAAGGAAGAAA 389
            :::::::::   ||||||                ::::::   ||
     105 ValSerGluLeuHisPheGlyAsnLeuAsnGlyAsnMetAsnAsnGlyAs 121

     390 TCCCGATCAACATCGTTTTACT.................TATAAAATTG 421
         |      :::||||||   :::                  |||      |
     121 nAlaLysSerHisArgAspValSerSerGluGluAsnArgTyrTyrThrV 138

     422 TGAAACGGAATAATTAT.........................AAAGCA 444
         ||::::::||||||:::                          |||
     138 alGluLysAsnAsnProThrGluAsnValThrSerPheThrLysGlu 154

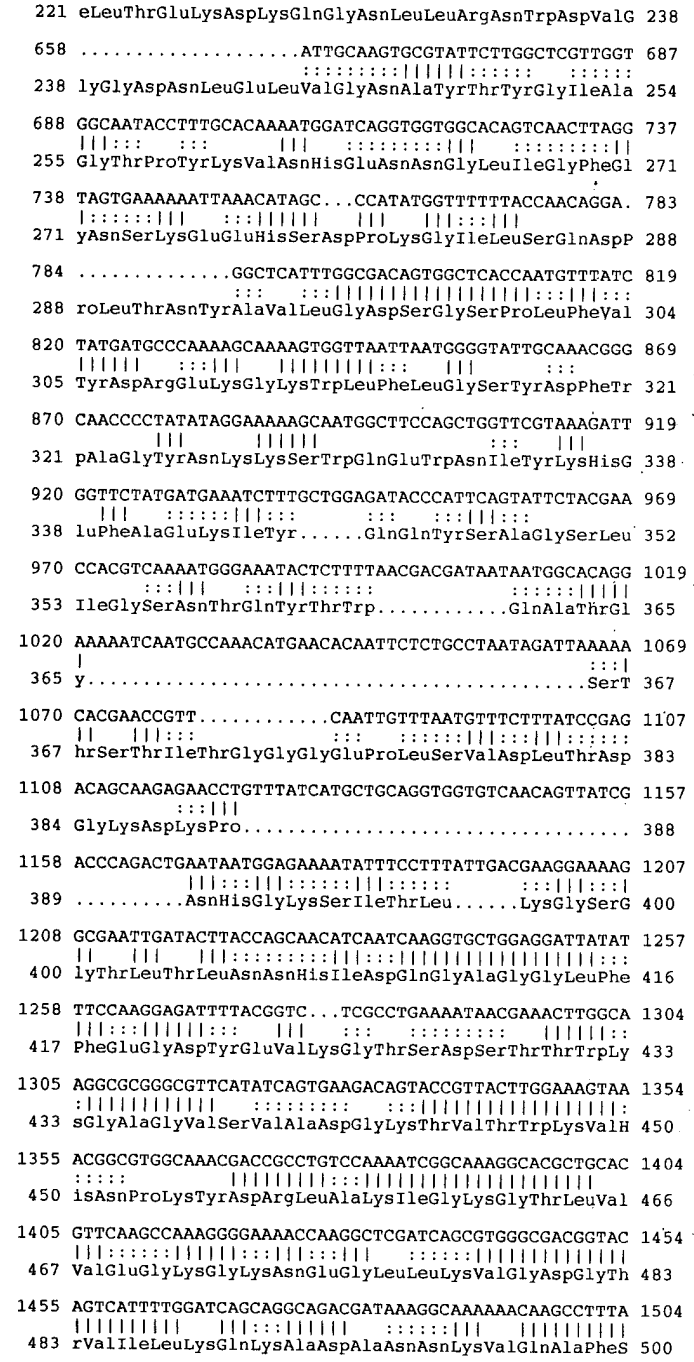
     445 GGGACTAAAGGCCATCCTTATGGCGGCGATTATCATATGCCGCGTTTGCA 494
                ::::::                |||||||:::|||||||||||||
     155 GluGlnAspAlaGlnLysArgArgGluAspTyrTyrMetProArgLeuAs 171

     495 TAAATTTGTCACAGATGCAGAACCTGTTGAAATGACCAGTTATATGGATG 544
          ||||||||||||::::::   |||:::|||   ::::::      ::::
     171 pLysPheValThrGluValAlaProIleGluAlaSerThrAlaAsnAsnA 188

     545 GGCGG...AAATATATCGATCAAAATAATTACCCTGACCGTGTTCGTATT 591
         :::::   :::|||   ::::::::::::|||||||      |||||||:::
     188 snLysGlyGluTyrAsnAsnSerAspLysTyrProAlaPheValArgLeu 204

     592 GGGGCAGGCAGGCAATAT............................TG 611
         |||:::|||   |||:::
     205 GlySerGlyThrGlnPheIleTyrLysLysGlySerArgTyrGlnLeuIl 221

     612 GCGATCTGATGAAGATGAGCCCAATAACCGCGAAAGTTCATATCAT.... 657
            :::::::::|||:::   :::|||         ::::::
```


221 eLeuThrGluLysAspLysGlnGlyAsnLeuLeuArgAsnTrpAspValG 238
658 ATTGCAAGTGCGTATTCTTGGCTCGTTGGT 687
238 lyGlyAspAsnLeuGluLeuValGlyAsnAlaTyrThrTyrGlyIleAla 254
688 GGCAATACCTTTGCACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGG 737
255 GlyThrProTyrLysValAsnHisGluAsnAsnGlyLeuIleGlyPheGl 271
738 TAGTGAAAAAATTAAACATAGC...CCATATGGTTTTTTACCAACAGGA. 783
271 yAsnSerLysGluGluHisSerAspProLysGlyIleLeuSerGlnAspP 288
784 GGCTCATTTGGCGACAGTGGCTCACCAATGTTTATC 819
288 roLeuThrAsnTyrAlaValLeuGlyAspSerGlySerProLeuPheVal 304
820 TATGATGCCCAAAAGCAAAAGTGGTTAATTAATGGGGTATTGCAAACGGG 869
305 TyrAspArgGluLysGlyLysTrpLeuPheLeuGlySerTyrAspPheTr 321
870 CAACCCCTATATAGGAAAAAGCAATGGCTTCCAGCTGGTTCGTAAAGATT 919
321 pAlaGlyTyrAsnLysLysSerTrpGlnGluTrpAsnIleTyrLysHisG 338
920 GGTTCTATGATGAAATCTTTGCTGGAGATACCCATTCAGTATTCTACGAA 969
338 luPheAlaGluLysIleTyr.....GlnGlnTyrSerAlaGlySerLeu 352
970 CCACGTCAAAATGGGAAATACTCTTTTAACGACGATAATAATGGCACAGG 1019
353 IleGlySerAsnThrGlnTyrThrTrp.....GlnAlaThrGl 365
1020 AAAAATCAATGCCAAACATGAACACAATTCTCTGCCTAATAGATTAAAAA 1069
365 y.....SerT 367
1070 CACGAACCGTT.....CAATTGTTTAATGTTTCTTTATCCGAG 1107
367 hrSerThrIleThrGlyGlyGlyGluProLeuSerValAspLeuThrAsp 383
1108 ACAGCAAGAGAACCTGTTTATCATGCTGCAGGTGGTGTCAACAGTTATCG 1157
384 GlyLysAspLysPro..... 388
1158 ACCCAGACTGAATAATGGAGAAAATATTTCCTTTATTGACGAAGGAAAAG 1207
389 AsnHisGlyLysSerIleThrLeu.....LysGlySerG 400
1208 GCGAATTGATACTTACCAGCAACATCAATCAAGGTGCTGGAGGATTATAT 1257
400 lyThrLeuThrLeuAsnAsnHisIleAspGlnGlyAlaGlyGlyLeuPhe 416
1258 TTCCAAGGAGATTTTACGGTC...TCGCCTGAAAATAACGAAACTTGGCA 1304
417 PheGluGlyAspTyrGluValLysGlyThrSerAspSerThrThrTrpLy 433
1305 AGGCGCGGGCGTTCATATCAGTGAAGACAGTACCGTTACTTGGAAAGTAA 1354
433 sGlyAlaGlyValSerValAlaAspGlyLysThrValThrTrpLysValH 450
1355 ACGGCGTGGCAAACGACCGCCTGTCCAAAATCGGCAAAGGCACGCTGCAC 1404
450 isAsnProLysTyrAspArgLeuAlaLysIleGlyLysGlyThrLeuVal 466
1405 GTTCAAGCCAAAGGGGAAAACCAAGGCTCGATCAGCGTGGGCGACGGTAC 1454
467 ValGluGlyLysGlyLysAsnGluGlyLeuLeuLysValGlyAspGlyTh 483
1455 AGTCATTTTGGATCAGCAGGCAGACGATAAAGGCAAAAAACAAGCCTTTA 1504
483 rValIleLeuLysGlnLysAlaAspAlaAsnAsnLysValGlnAlaPheS 500
1505 GTGAAATCGGCTTGGTCAGCGGCAGGGGTACGGTGCAACTGAATGCCGAT 1554
500 erGlnValGlyIleValSerGlyArgSerThrLeuValLeuAsnAspAsp 516
1555 AATCAGTTCAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTT 1604
517 LysGlnValAspProAsnSerIleTyrPheGlyPheArgGlyGlyArgLe 533
1605 GGATTTAAACGGGCATTCGCTTTCGTTCCACCGTATTCAAAATACCGATG 1654
533 uAspLeuAsnGlyAsnSerLeuThrPheAspHisIleArgAsnIleAspA 550
1655 AAGGGGCGATGATTGTCAACCACAATCAAGACAAAGAATCCACCGTTACC 1704
550 spGlyAlaArgValValAsnHisAsnMetThrAsnThrSerAsnIleThr 566
1705 ATTACAGGC..... 1713
567 IleThrGlyGluSerLeuIleThrAsnProAsnThrIleThrSerTyrAs 583
1713 1713
583 nIleGluAlaGlnAspAspAspHisProLeuArgIleArgSerIleProT 600
1713 1713
600 yrArgGlnLeuTyrPheAsnGlnAspAsnArgSerTyrTyrThrLeuLys 616
1713 1713
617 LysGlyAlaSerThrArgSerGluLeuProGlnAsnSerGlyGluSerAs 633
1714 AATAAAGATATTGCTACAACCGGCA 1738
633 nGluAsnTrpLeuTyrMetGlyArgThrSerAspAlaAlaLysArgAsnV 650
1739 ATAACAACAGCTTGGATAGCAAAAAAGAAATTGCCTACAACGGTTGGTTT 1788
650 alMetAsnHisIleAsnAsnGluArgMetAsnGlyPheAsnGlyTyrPhe 666
1789 GGCGAGAAAGATACGACCAAAACG...AACGGGCGGCTCAACCTTGTTTA 1835
667 GlyGluGluGluThrLysAlaThrGlnAsnGlyLysLeuAsnValThrPh 683
1836 CCAGCCCGCCGCAGAAGACCGCACCCTGCTGCTTTCCGGCGGAACAAATT 1885
683 eAsnGlyLysSerAspGlnAsnArgPheLeuLeuThrGlyGlyThrAsnL 700
1886 TAAACGGCAACATCACGCAAACAAACGGCAAACTGTTTTTCAGCGGCAGA 1935
700 euAsnGlyAspLeuAsnValGluLysGlyThrLeuPheLeuSerGlyArg 716
1936 CCAACACCGCACGCCTACAATCATTTAAACGACCATTGGTCGCAAAAAGA 1985
717 ProThrProHisAlaArgAspIleAlaGlyIleSerSerThrLysLysAs 733
1986 GGGC.....ATTCCTCGCGGGGAAATCGTGTGGGACAACGACTGGATCA 2029
733 pProHisPheThrGluAsnAsnGluValValValGluAspAspTrpIleA 750
2030 ACCGCACATTTAAAGCGGAAAACTTCCAAATTAAAGGCGGACAGGCGGTG 2079
750 snArgAsnPheLysAlaThrThrMetAsnValThrGlyAsnAlaSerLeu 766
2080 GTTTCC...CGCAATGTTGCCAAAGTGAAAGGCGATTGGCATTTGAGCAA 2126
767 TyrSerGlyArgAsnValAlaAsnIleThrSerAsnIleThrAlaSerAs 783
2127 TCACGCCCAAGCAGTTTTTGGTGTCGCACCGCATCAAAGCCACACAATCT 2176
783 nAsnAlaGlnValHisIleGlyTyr.....LysThrGlyAspThrValC 798



```
1368 LysProAlaAlaLysProGlnAlaGlnAlaLysProGlnThrGluProAl 1384

3327 CGGGGAAAATGTCGGCATTATGCAGGCGGAGGAA..................... 3360
     :   |||||||||:::   ::::::::::::|||
1384 aArgGluAsnValSerThrValAsnThrLysGluProGlnSerGlnThrS 1401

3360 ..................................................... 3360

1401 erAlaThrValSerThrGluGlnProAlaLysGluThrSerSerAsnVal 1417

3360 ..................................................... 3360

1418 GluGlnProAlaProGluAsnSerIleAsnThrGlySerAlaThrThrMe 1434

3360 ..................................................... 3360

1434 tThrGluThrAlaGluLysSerAspLysProGlnMetGluThrValThrG 1451

3360 ..................................................... 3360

1451 luAsnAspArgGlnProGluAlaAsnThrValAlaAspAsnSerValAla 1467

3360 ..................................................... 3360

1468 AsnAsnSerGluSerSerGluSerLysSerArgArgArgArgSerValSe 1484

3361 .GAGAAAAAACGGGTGCAGGCGGATAAAGACACCGCCTTGGCGAAACAGC 3409
      :::   |||:::::::::|||::::::   |||:::   :::   |||:
1484 rGlnProLysGluThrSerAlaGluGluThrThrValAlaSerThrGlnG 1501

3410 GCGAAGCGGAAACCCGGCCGGCTACCACCGCCTTCCCCCGCGCCCGCCGC 3459
     ::   :::      :::   ::::::|||      ||||||:::||||||
1501 luThrThrValAspAsnSerValSerThrProLysProArgSerArgArg 1517

3460 GCCCGCCGGGATTTGCCGCAACTGCAACCCCAACCGCAGCCCCAACCGCA 3509
     :::||||||::::::        :::   :::|||         |||
1518 ThrArgArgSerValGlnThrAsnSerTyrGluProValGluLeuProTh 1534

3510 G..................................................... 3510

1534 rGluAsnAlaGluAsnAlaGluAsnValGlnSerGlyAsnAsnValAlaA 1551

3511 .................CGCGACCTGATCAGCCGTTATGCCAATAGCGGT 3543
                     |||:::|||   |||:::   :::|||:::
1551 snSerGlnProAlaLeuArgAsnLeuThrSerLysAsnThrAsnAlaVal 1567

3544 TTGAGTGAATTTTCCGCC..............ACGCTCAACAGCGTTTT 3578
     :::|||:::      |||              :::|||||
1568 IleSerAsnAlaMetAlaLysAlaGlnPheValAlaLeuAsnValGlyLy 1584

3579 CGCCGTACAGGACGAATTAGACCGCGTA...TTTGCCGAAGACCGCCGCA 3625
      ||||||:::::::::::::::::::::   :::   ::::::   :::
1584 sAlaValSerGlnHisIleSerGlnLeuGluMetAsnAsnGluGlyGlnT 1601

3626 ACGCCGTTTGGACAAGCGGCATCCGGGACACCAAACACTACCGTTCGCAA 3675
          ||||||   |||:::          :::|||:::|||   |||:::
1601 yrAsnValTrpIleSerAsnThrSerMetAsnLysAsnTyrSerSerGlu 1617

3676 GATTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCAAATCGGTATGCA 3725
     :::::::|||   :::   :::::::::         |||:::|||   ::
1618 GlnTyrArgArgPheSerSerLysSerThrGlnThrGlnLeuGlyTrpAs 1634

3726 GAAAAACCTCGGCAGCGGG...CGCGTCGGCATCCTGTTTTCGCACAACC 3772
     ::::::::::::::::::   ::::::|||   :::|||::::::   |
1634 pGlnThrIleSerAsnAsnValGlnLeuGlyGlyValPheThrTyrValA 1651

3773 GGACCGAAAACACCTTCGACGACGGCATCGGCAACTCGGCACGGCTTGCC 3822
     ||::::::|||:::||||||   :::   ::::::::::::   ||||||
1651 rgAsnSerAsnAsnPheAspLysAlaSerSerLysAsnThr...LeuAla 1666

3823 CACGGCGCCGTTTTCGGGCAATACGGCATCGAC...AGGTTCTACATCGG 3869
     :::         ::::::::::|||      |||   ::::::|||:::||
1667 GlnValAsnPheTyrSerLysTyrTyrAlaAspAsnHisTrpTyrLeuGl 1683

3870 CATCAGCGCGGGGCGCGGGTTTTAGCAGCGGCAGCCTTTCAGACGGCATCG 3919
     ||||:::   |||   |||      :::::::::|||:::   :::   :
1683 yIleAspLeuGlyTyrGlyLysPheGlnSerAsnLeuGlnThrAsnAsnA 1700

3920 GAGGCAAAATCCGCCGCCGCGTGCTGCATTACGGCATTCAGGCACGATAC 3969
     :::::|||:::   |||::::::   :::   |||:::   |||
1700 snAlaLysPheAlaArgHisThrAlaGlnIleGlyLeuThrAlaGlyLys 1716

3970 CGCGCCGGTTTCGGCGGATTCGGCATCGAACCGCACATCGGCGCAACGCG 4019
          ::::::|||:::|||:::::::::|||   :::|||:::
1717 AlaPheAsnLeuGlyAsnPheAlaValLysProThrValGlyValArgTy 1733

4020 CTATTTCGTCCAAAAAGCGGATTACCGCTACGAAAACGTCAATATCGCCA 4069
          ::::::::::::||||||:::        :::   :::|||   :
1733 rSerTyrLeuSerAsnAlaAspPheAlaLeuAlaGlnAspArgIleLysV 1750

4070 CCCCCGGCCTTGCATTCAACCGCTACCGCGCGGGCATTAAGGCAGATTAT 4119
     ::      ::::::   :::         |||   :::      :::|||
1750 alAsnProIleSerValLysThrAlaPheAlaGlnValAspLeuSerTyr 1766

4120 TCATTCAAACCGGCGCAACACATTTCCATCACGCCTTATTTGAGCCTGTC 4169
     ::::::      ::::::   :::||||||||||||   ||||||
1767 ThrTyrHisLeuGlyGlu...PheSerIleThrProIleLeuSerAlaAr 1782

4170 CTATACCGATGCCGCTTCGGGCAAAGTCCGAACACGCGTCAATACCGCCG 4219
      |||   ||||||   :::|||:::   :::   :::|||::::::
1782 gTyr...AspAlaAsnGlnGlyAsnGlyLysIleAsnValSerValTyrA 1798

4220 TATTGGCTCAGGATTTCGGCAAAACCCGCAGTGCGGAATGGGGCGTAAAC 4269
       :::|||   :::      :::   ::::::   :::   |||::::::
1798 spPheAlaTyrAsnValGluAsnGlnGlnGlnTyrAsnAlaGlyLeuLys 1814

4270 GCCGAAATCAAAGGTTTCACGCTGTCCCTCCACGCTGCCGCCGCCAAAGG 4319
       :::      :::      |||||||||   ::::::   :::|||::
1815 LeuLysTyrHisAsnValLysLeuSerLeuIleGlyGlyLeuThrLysAl 1831

4320 CCCGCAACTGGAAGCGCAACACAGCGCGGGCATCAAATTAGGCTAC 4365
     :   |||   |||   |||   :::|||   :::||||||::::::
1831 aLysGlnAlaGluLysGlnLysThrAlaGluValLysLeuSerPhe 1846
```

seq_name: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-460-269C-4

```
seq_documentation_block:
; Sequence 4, Application US/08460269C
; Patent No. 6197548
;   GENERAL INFORMATION:
;        APPLICANT: CLARE, JEFFREY J.
;                   ROMANOS, MICHAEL A.
;        TITLE OF INVENTION: EXPRESSION OF HETEROLOGOUS PROTEIN IN
;                            YEAST
;        NUMBER OF SEQUENCES: 17
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Millen, White, Zelano & Branigan, P.C.
;             STREET: 2200 Clarendon Blvd., Suite 1400
;             CITY: ARLINGTON
;             STATE: VA
;             COUNTRY: USA
;             ZIP: 22201
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/08/460,269C
;             FILING DATE: 02-Jun-1995
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Lebovitz, Richard M.
;             REGISTRATION NUMBER: 37,067
;             REFERENCE/DOCKET NUMBER: Popov-2
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (703) 243-6333
```

```
;            TELEFAX: (703) 243-6410
;   INFORMATION FOR SEQ ID NO: 4:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 911 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-08-460-269C-4


alignment_scores:
            Quality:  295.50              Length:   1079
              Ratio:  0.612                 Gaps:     50
 Percent Similarity:  44.764    Percent Identity: 21.965

alignment_block:
 US-09-303-518D-649 x US-08-460-269C-4   ..

 Align seg 1/1  to: US-08-460-269C-4  from: 1  to: 911

    1417 GGGGAAAACCAAGGCTCGATCAGCGTGGGCGACGGTACAGTCATTTTGGA 1466
         ||||||:::|||   :::|||   :::
      45 GlyGluArgGlnHisGlyIleHisIle..................... 53

    1467 TCAGCAGGCAGACGATAAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCT 1516
          :::|||:::|||      |||   :::   |||   :::
      54 .LysGlnSerAspGlyAlaGlyValArgThrAlaThrGlyThrThrIleL 70

    1517 TGGTCAGCGGCAGGGGTACGGTGCAACTGAATGCCGATAATCAGTTCAAC 1566
           ||||||||||||   :::      :::      :::|||
      70 ysValSerGlyArgGlnAlaGlnGlyValLeuLeuGluAsn......... 83

    1567 CCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAACGG 1616
                    :::   |||:::::::|||   :::      :::||
      84 ...ProAlaAlaGluLeuArgPheGlnAsnGlySerValThrSerSerGl 99

    1617 GCATTCGCTTTCGTTCCACCGTATTCAAAATACCGATGAAGGGGCGATGA 1666
         |                       |||      |||||||||:::
      99 y.......................GlnLeuPheAspGluGlyVal.... 106

    1667 TTGTCAACCACAATCAAGACAAAGAATCCACCGTTACCATTACAGGCAAT 1716
                    :::         :::||||||||||:::   ::::::
     107 ..............ArgArgPheLeuGlyThrValThrValLysAlaGly 118

    1717 AAA.........GATATTGCTACAACCGGCAATAACAACAGCTTGGATAG 1757
         |||         |||   ||||||   :::|||
     119 LysLeuValAlaAspHisAlaThrLeuAlaAsn................. 129

    1758 CAAAAAAGAAATTGCCTACAACGGTTGGTTTGGCGAGAAAGATACGACCA 1807
                                              :::||||||
     130 ................................ValSerAspThrArgA 135

    1808 AAACGAACGGGCGGCTCAACCTTGTT............TACCAGCCCGCC 1845
          :::|||            |||              |||   :::
     135 spAspAspGlyIleAlaLeuTyrValAlaGlyGluGlnAlaGlnAlaSer 151

    1846 GCAGAAGACCGCACCCTGCTGCTTTCCGGCGGAACAAATTTA......AA 1889
               |||   ||||||      :::||||||:::::::::      ::
     152 IleAlaAspSerThrLeuGlnGlyAlaGlyGlyValArgValGluArgGl 168

    1890 CGGCAACATCACGCAAACAAACGGCAAACTGTTTTTCAGCGGC....... 1932
         ::::|||:::|||      ::::::   :::      :::|||
     168 yAlaAsnValThrValGlnArgSerThrIleValAspGlyGlyLeuHisI 185

    1933 ...........AGACCAACACCGCACGCCTACAATCATTTAAACGACCAT 1971
                     :::|||
     185 leGlyThrLeuGlnProLeu............................. 191

    1972 TGGTCGCAAAAAGAGGGCATTCCTCGCGGGGAAATCGTGTGGGACAACGA 2021
               |||   |||   :::|||   ::::::::::|||      :::
     192 ......GlnProGluAspLeuProProSerArgValValLeuGlyAspTh 206

    2022 CTGGATCAAC.............................CGCACAT 2038
               ::::::                           :::|
     206 rSerValThrAlaValProAlaSerGlyAlaProAlaAlaValSerValP 223

    2039 TTAAAGCGGAAAACTTCCAAATTAAAGGCGGACAGGCGGTGGTTTCCCGC 2088
         ||   |||:::::::::   :::   ||||||:::   :::   :::|||
     223 heGlyAlaAsnGluLeuThrValAspGlyGlyHisIleThrGlyGlyArg 239

    2089 AATGTTGCCAAAGTGAAAGGCGATTGGCATTTGAGCAATCACGCCCAAGC 2138
               ::::::                                 :::   ||
     240 AlaAlaGly.............................ValAlaAl 245

    2139 AGTTTTTGGTGTCGCACCGCATCAAAGCCACACAATCTGTACACGTTCGG 2188
         |:::   |||:::      |||   :::::::::         |||:::|
     245 aMetAspGlyAlaIleValHisLeuGlnArgAlaThrIleArgArgGlyA 262

    2189 AC.......................................... 2190
         ||
     262 spAlaProAlaGlyGlyAlaValProGlyGlyAlaValProGlyGlyPhe 278

    2191 .................TGGACGGGTCTGACAAATTGTGTCGAAAAAAC 2222
                         |||   |||:::
     279 GlyProLeuLeuAspGlyTrpTyrGlyVal..................... 288

    2223 CATTACCGACGATAAAGTGATTGCTTCATTGACTAAGACCGACATCAGC. 2271
                                                 |||:::|||
     289 .....................................AspValSerA 292

    2272 ..GGCAATGTCGATCTTGCCGATCACGCTCATTTAAATCTCACAGGGCTT 2319
           ::::::|||||||||||||:::           :::::::::
     292 spSerThrValAspLeuAlaGln............SerIleValGluAla 304

    2320 GCCACACTCAACGGCAATCTTAGTGCAAATGGCGATACACGTTATACAGT 2369
               |||::::::   :::   |||:::      :::|||   |||||
     305 ProGlnLeuGlyAlaAlaIleArgAlaGlyArgGlyAlaArgValThrVa 321

    2370 CAGCCACAAC......GCCACCCAAAACGGCAACCTTAGCCTCGTGGGCA 2413
         ||||   :::      ::::::   :::||||||:::      :::|||:
     321 lSerGlyGlySerLeuSerAlaProHisGlyAsnValIleGluThrGlyG 338

    2414 ATGCCCAAGCAACATTTAAT.................CAAGCCACATTA 2445
         :::::        |||                    :::   ||||||
     338 lyGlyAlaArgArgProProProAlaSerProLeuSerIleThrLeu 354

    2446 AACGGCAACACATCGGCTTCGGGCAATGCT...TCATTTAATCTAAGCGA 2492
         ::::::::::   |||:::|||:::|||      :::::::::
     355 GlnAlaGlyAlaArgAlaGlnGlyArgAlaLeuLeuTyrArgValLeuPr 371

    2493 CCACGCCGTACAAAACGGCAGTCTGACGCTTTCCGGCAACGCTAAGGCAA 2542
          :::   |||:::         |||||||||:::|||:::|||::::::
     371 oGluProValLys.........LeuThrLeuAlaGlyGlyAlaGlnGly. 384

    2543 ACGTAAGCCATTCCGCACTCAACGGTAATGTCTCCCTAGCCGATAAGGCA 2592
                           :::|||::::::      :::
     385 ...................GlnGlyAspIleValAlaThr......... 391

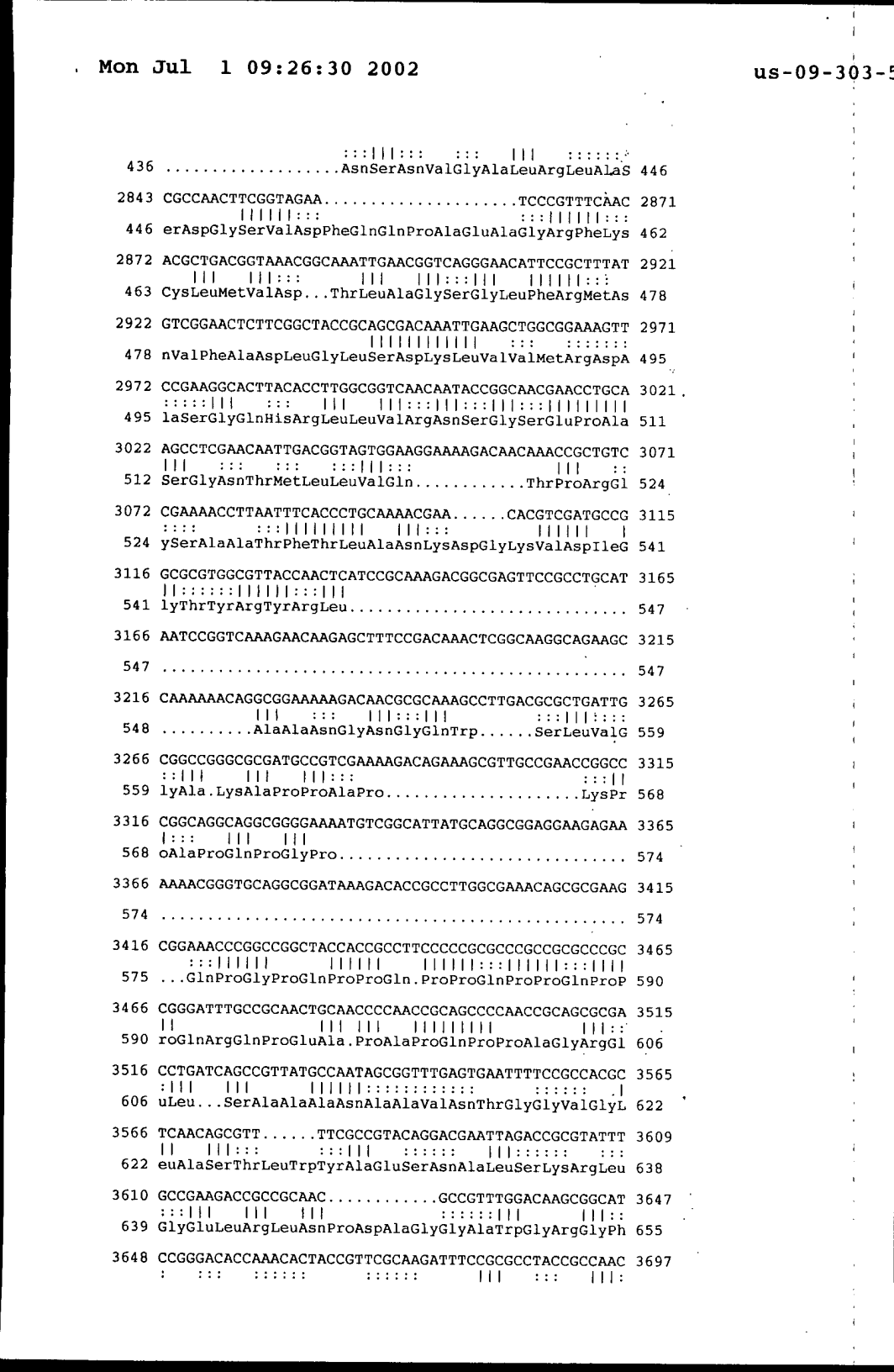
    2593 GTATTCCATTTTGAAAGCAGCCGCTTTACCGGACAAATCAGCGGCGGCAA 2642
                    |||          :::   |||      ||||||
     392 ............GluLeuProProIleProGlyAlaSerSerGlyProLe 404

    2643 GGATACGGCATTACACTTAAAAGACAGCGAATGGACGCTGCCGTCAGGCA 2692
          |||:::||||||      :::::::::::::||||||      :::
     404 uAspValAlaLeu...AlaSerGlnAlaArgTrpThr...GlyAlaThrA 419

    2693 CGGAATTAGGCAATTTAAACCTTGACAACGCCACCATTACACTCAATTCC 2742
               :::   :::|||::::::||||||||||||   ::::::
     419 rgAlaValAspSerLeuSerIleAspAsnAlaThrTrpValMet...... 433

    2743 GCCTATCGCCACGATGCGGCAGGGGCGCAAACCGGCAGTGCGACAGATGC 2792
                                                  ||||||
     434 ......................................ThrAsp.. 435

    2793 GCCGCGCCGCCGTTCGCGCCGTTCGCGCCGTTCCCTATTATCCGTTACAC 2842
```



seq_name: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-728-470-10

seq_documentation_block:
; Sequence 10, Application US/08728470
; Patent No. 5928651
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:

```
;      ADDRESSEE:  Shoemaker and Mattare, Ltd.
;      STREET:  2001 Jefferson Davis Hwy., 1203 Crystal Plaza
;      STREET:  Bldg. 1
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202-0286
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/728,470
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/302,832
;      FILING DATE:  16-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US PCT/US93/02166
;      FILING DATE:  16-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9205704.1
;      FILING DATE:  16-MAR-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Berkstresser, Jerry W
;      REGISTRATION NUMBER:  22,651
;      REFERENCE/DOCKET NUMBER:  1038-633
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 415-0810
;      TELEFAX:  (703) 415-0813
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1529 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-728-470-10


alignment_scores:
            Quality:  295.50            Length:   1196
              Ratio:  0.505               Gaps:     62
 Percent Similarity:  48.913   Percent Identity: 20.819

alignment_block:
 US-09-303-518D-649 x US-08-728-470-10   ..

 Align seg 1/1  to: US-08-728-470-10  from: 1  to: 1529

     289 TCGCGTAACGGCGTGGCGGCATTGGTGGGCGATCAATATATTGTGAGCGT 338
         :::||||||::::::   |||   :::      |||      ::::::||
     507 AlaArgAsnAlaThrAspAlaGlnIleThrAlaGlnGlyThrIleThrVa 523

     339 GGCACATAACGGC.........GGCTATAACAACGTTGATTTTGGTGCGG 379
         |      :::              :::||||||||||:::::::::::::
     523 lAsnLysAspAspLysGlnPheArgPheAsnAsnValSerIleAsnGlyT 540

     380 AAGGAAGAAATCCCGATCAACATCGTTTTACTTATAAAATTGTGAAACGG 429
         |||::::::            :::|||         |||:::::::::
     540 hrGlyLysGly.........LeuLysPhe.........IleAlaAsnGln 550

     430 AATAATTATAAAGCAGGGACTAAAGGCCATCCTTATGGCGGCGATTATCA 479
         ||||||:::                |||   :::   |||::  , ::
     551 AsnAsnPheThr..............HisLysPheAspGlyGluLeuAs 562

     480 TATGCCGCGTTTGCATAAATTTGTCACAGATGCAGAACCTGTTGAAATGA 529
         :
     562 n..................................................  562

     530 CCAGTTATATGGATGGGCGGAAATATATCGATCAAAATAATTACCCTGAC 579
                 ::::::|||         |||:::|||::::::      |||

     563 ........IleSerGlyIleValThrIleAsnGlnThrThrLysLysAsp 576

     580 CGTGTTCGTATTGGGGCAGGCAGGCAATATTGGCGATCTGATGAAGATGA 629
                                  :::||||||:::::::::::::|||::
     577 ....................ValLysTyrTrpAsnAlaSerLysAspSe 586

     630 GCCCAATAACCGCGAAAGT...TCATATCATATTGCAAGTGCGTATTCTT 676
         :      |||   :::|||   :::   :::   ::::::   :::::::
     586 rTyrTrpAsnValSerSerLeuThrLeuAsnThrValGlnLysPheThrP 603

     677 GGCTC............GTTGGTGGCAAT..................... 693
         :::::              |||:::|||
     603 heIleLysPheValAspSerGlySerAsnSerGlnAspLeuArgSerSer 619

     694 ......ACCTTTGCACAA.........AATGGATCAGGTGGTGGCACA.. 726
               :::|||||||           ||||||   ||||||   |||
     620 ArgArgSerPheAlaGlyValHisPheAsnGlyIleGlyGlyLysThrAs 636

     727 .GTCAACTTAGGTAGTGAA..............AAAATTAAACATAGC. 759
         |||:::|||::::::                |||:::|||   :::
     636 nPheAsnIleGlyAlaAsnAlaLysAlaLeuPheLysLeuLysProAsnA 653

     760 ...........CCATATGGTTTTTTACCA..................... 777
                    |||          ||||||
     653 laAlaThrAspProLysLysGluLeuProIleThrPheAsnAlaAsnIle 669

     778 ......ACAGGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATCTA 821
               ||||||:::|||          :::::::|||   ||||||
     670 ThrAlaThrGlyAsnSer.........AspSerSerValMetPheAspIl 683

     822 TGATGCCCAAAAGCAAAAGTGGTTA............ATTAATGGGGTAT 859
         |||:::      :::                      ::::::::::::
     683 eHisAlaAsnLeuThrSerArgAlaAlaGlyIleAsnMetAspSerIleA 700

     860 TGCAAACGGGC..................AACCCCTATATAGGAAAAAGC 891
               ||||||                 :::   :::      :::|||
     700 snIleThrGlyGlyLeuAspPheSerIleThrSerHisAsnArgAsnSer 716

     892 AATGGCTTCCAGCTGGTTCGT........................... 912
         |||:::|||::::::   :::
     717 AsnAlaPheGluIleLysLysAspLeuThrIleAsnAlaThrGlySerAs 733

     913 ...................AAAGATTGGTTCTATGATGAAATCTTTGCTG 943
                             ||||||   ||||||:::|||
     733 nPheSerLeuLysGlnThrLysAspSerPheTyrAsnGluTyrSerLysH 750

     944 GAGATACCCATTCAGTATTCTACGAACCACGTCAAAATGGGAAATACTCT 993
                   :::|||                  :::|||:::   :::
     750 isAlaIleAsnSerSerHisAsnLeuThrIleLeuGlyGlyAsnValThr 766

     994 TTTAACGACGATAATAATGGC.........ACAGGAAAAATCAATGCCAA 1034
         ::::::   :::|||::::::         ||||||:::||||||
     767 LeuGlyGlyGluAsnSerSerSerSerIleThrGlyAsnIleAsnIleTh 783

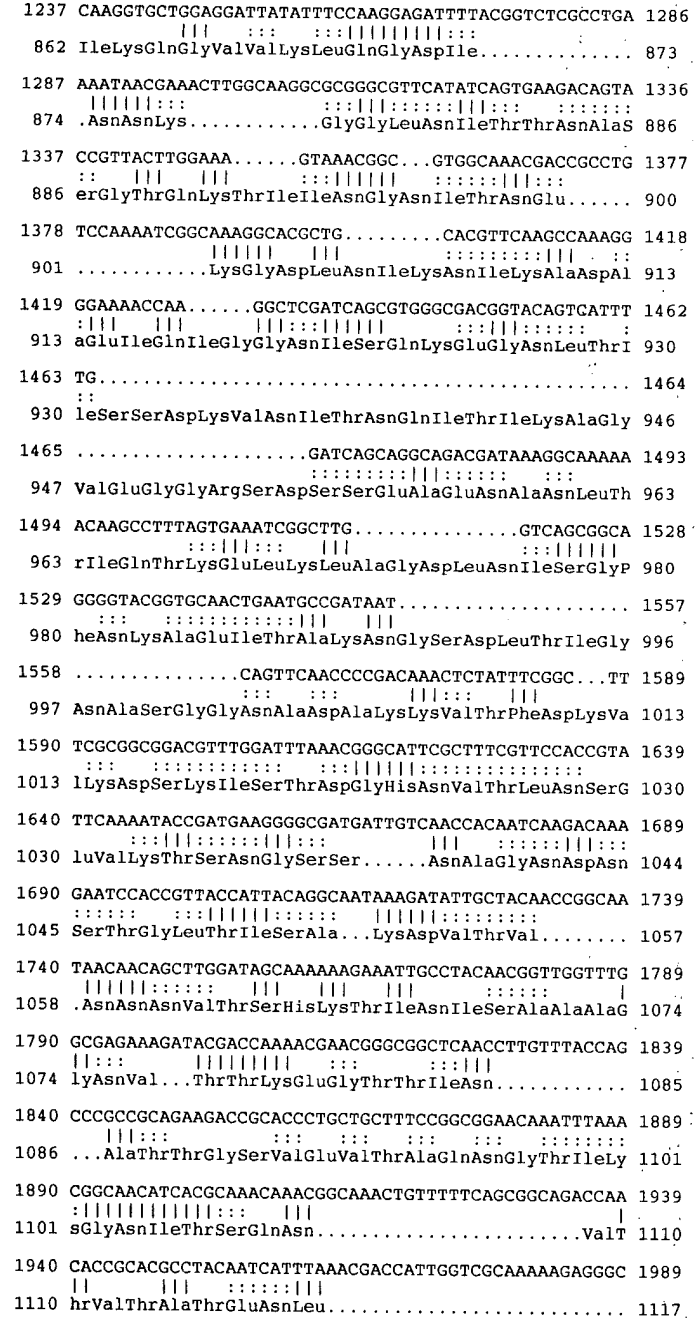
    1035 ACATGAACACAAT.......................TCTCTGCCTAATA 1060
          ::::::   |||                       :::      :::
     783 rAsnLysAlaAsnValThrLeuGlnAlaAspThrSerAsnSerAsnThrG 800

    1061 GATTAAAAACACGAACCGTTCAATTGTTTAATGTTTCTTTATCCGAGACA 1110
           ||||||   ||||||:::   |||   |||:::|||:::
     800 lyLeuLysLysArgThrLeuThrLeuGlyAsnIleSerVal......... 813

    1111 GCAAGAGAACCTGTTTATCATGCTGCAGGTGGTGTCAACAGTTATCGACC 1160
               |||                ::::|||:::      :::
     814 ......GluGlyAsnLeuSerLeuThrGlyAlaAsnAlaAsnIleValGl 828

    1161 CAGACTGAATAATGGAGAAAATATTTCCTTTATTGACGAAGGAAAAGGCG 1210
          :::|||:::   :::|||:::   :::|||      |||::::::   :
     828 yAsnLeuSerIleAlaGluAspSerThrPheLysGlyGluAlaSerAspA 845

    1211 AATTGATACTT.......................ACCAGCAACATCAAT 1236
         ::|||   :::                       |||:::|||||||||
     845 snLeuAsnIleThrGlyThrPheThrAsnAsnGlyThrAlaAsnIleAsn 861
```


1237 CAAGGTGCTGGAGGATTATATTTCCAAGGAGATTTTACGGTCTCGCCTGA 1286
862 IleLysGlnGlyValValLysLeuGlnGlyAspIle..... 873
1287 AAATAACGAAACTTGGCAAGGCGCGGGCGTTCATATCAGTGAAGACAGTA 1336
874 .AsnAsnLys.....GlyGlyLeuAsnIleThrThrAsnAlaS 886
1337 CCGTTACTTGGAAA......GTAAACGGC...GTGGCAAACGACCGCCTG 1377
886 erGlyThrGlnLysThrIleIleAsnGlyAsnIleThrAsnGlu..... 900
1378 TCCAAAATCGGCAAAGGCACGCTG.....CACGTTCAAGCCAAAGG 1418
901 LysGlyAspLeuAsnIleLysAsnIleLysAlaAspAl 913
1419 GGAAAACCAA......GGCTCGATCAGCGTGGGCGACGGTACAGTCATTT 1462
913 aGluIleGlnIleGlyGlyAsnIleSerGlnLysGluGlyAsnLeuThrI 930
1463 TG..... 1464
930 leSerSerAspLysValAsnIleThrAsnGlnIleThrIleLysAlaGly 946
1465 GATCAGCAGGCAGACGATAAAGGCAAAAA 1493
947 ValGluGlyGlyArgSerAspSerSerGluAlaGluAsnAlaAsnLeuTh 963
1494 ACAAGCCTTTAGTGAAATCGGCTTG.....GTCAGCGGCA 1528
963 rIleGlnThrLysGluLeuLysLeuAlaGlyAspLeuAsnIleSerGlyP 980
1529 GGGGTACGGTGCAACTGAATGCCGATAAT..... 1557
980 heAsnLysAlaGluIleThrAlaLysAsnGlySerAspLeuThrIleGly 996
1558 CAGTTCAACCCCGACAAACTCTATTTCGGC...TT 1589
997 AsnAlaSerGlyGlyAsnAlaAspAlaLysLysValThrPheAspLysVa 1013
1590 TCGCGGCGGACGTTTGGATTTAAACGGGCATTCGCTTTCGTTCCACCGTA 1639
1013 lLysAspSerLysIleSerThrAspGlyHisAsnValThrLeuAsnSerG 1030
1640 TTCAAAATACCGATGAAGGGGCGATGATTGTCAACCACAATCAAGACAAA 1689
1030 luValLysThrSerAsnGlySerSer.....AsnAlaGlyAsnAspAsn 1044
1690 GAATCCACCGTTACCATTACAGGCAATAAAGATATTGCTACAACCGGCAA 1739
1045 SerThrGlyLeuThrIleSerAla...LysAspValThrVal..... 1057
1740 TAACAACAGCTTGGATAGCAAAAAAGAAATTGCCTACAACGGTTGGTTTG 1789
1058 .AsnAsnAsnValThrSerHisLysThrIleAsnIleSerAlaAlaAlaG 1074
1790 GCGAGAAAGATACGACCAAAACGAACGGGCGGCTCAACCTTGTTTACCAG 1839
1074 lyAsnVal...ThrThrLysGluGlyThrThrIleAsn..... 1085
1840 CCCGCCGCAGAAGACCGCACCCTGCTGCTTTCCGGCGGAACAAATTTAAA 1889
1086 ...AlaThrThrGlySerValGluValThrAlaGlnAsnGlyThrIleLy 1101
1890 CGGCAACATCACGCAAACAAACGGCAAACTGTTTTTCAGCGGCAGACCAA 1939
1101 sGlyAsnIleThrSerGlnAsn.....ValT 1110
1940 CACCGCACGCCTACAATCATTTAAACGACCATTGGTCGCAAAAAGAGGGC 1989
1110 hrValThrAlaThrGluAsnLeu..... 1117


1990 ATTCCTCGCGGGGAAATCGTGTGGGACAACGACTGGATCAACCGCACATT 2039
1118 ValThrThrGluAsnAlaValIleAsnAlaThrSe 1129
2040 TAAAGCGGAAAACTTCCAAATTAAAGGCGGACAGGCGGTGGTTTCCCGCA 2089
1129 rGlyThrValAsnIleSerThrLysThrGlyAsp.....IleLysGlyG 1144
2090 ATGTTGCCAAAGTGAAAGGCGATTGGCATTTGAGCAATCACGCCCAAGCA 2139
1144 lyIleGluSerThrSerGlyAsnValAsnIleThrAlaSerGlyAsnThr 1160
2140 GTTTTTGGTGTCGCACCGCATCAAAGCCACACAATCTGTACACGTTCGGA 2189
1161 Leu...LysValSerAsnIleThrGlyGlnAspValThrValThrAlaAs 1176
2190 CTGGACGGGTCTGACAAATTGTGTCGAAAAAACCATTACCGACGATAAAG 2239
1176 pAlaGlyAlaLeuThrThrThrAlaGlySerThrIleSerAlaThrThrG 1193
2240 TGATTGCTTCATTG...ACTAAGACC...GACATCAGCGGCAATGTCGAT 2283
1193 lyAsnAlaAsnIleThrThrLysThrGlyAspIleAsnGlyLysValGlu 1209
2284 CTTGCCGAT...CACGCTCATTTAAATCTCACAGGGCTTGCCACACTCAA 2330
1210 SerSerSerGlySerValThrLeuValAlaThrGlyAlaThrLeuAlaVa 1226
2331 CGGCAATCTTAGTGCAAATGGCGATACA.....CGTT 2362
1226 lGlyAsnIleSerGlyAsnThrValThrIleThrAlaAspSerGlyLysL 1243
2363 ATACAGTCAGCCACAACGCCACCCAAAACGGCAACCTTAGCCTCGTGGGC 2412
1243 euThrSerThrValGlySerThrIleAsnGlyThrAsnSerValThrThr 1259
2413 AATGCCCAAGCAACATTTAATCAAGCCACATTAAACGGCAACACATCGGC 2462
1260 SerSerGlnSerGlyAspIleGluGlyThrIleSerGlyAsnThr..... 1274
2463 TTCGGGCAATGCTTCATTTAATCTAAGCGACCACGCCGTACAAAACGGCA 2512
1275 ValAsnValThr.....AlaSerThrGlyA 1283
2513 GTCTGACGCTTTCCGGCAACGCTAAGGCAAACGTAAGCCATTCCGCACTC 2562
1283 spLeuThrIleGlyAsnSerAlaLysValGluAlaLysAsnGlyAla... 1298
2563 AACGGTAATGTCTCCCTAGCCGATAAGGCAGTATTCCATTTTGAAAGCAG 2612
1299 AlaThrLeuThrAlaGluSerGl 1306
2613 CCGCTTTACCGGACAA.....ATCAGCGGCGGCAAGGATA 2647
1306 yLysLeuThrThrGlnThrGlySerSerIleThrSerSerAsnGlyGlnT 1323
2648 CGGCATTACACTTAAAAGACAGCGAATGGACGCTGCCGTCAGGCACGGAA 2697
1323 hrThrLeuThrAlaLysAspSer.....SerIle 1332
2698 TTAGGCAATTTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTA 2747
1333 AlaGlyAsnIleAsnAla.....AlaAsnValThrLeuAsnThr..... 1345
2748 TCGCCACGATGCGGCAGGGGCGCAAACCGGCAGTGCGACAGATGCGCCGC 2797
1346 ThrGlyThr..... 1348
2798 GCCGCCGTTCGCGCCGTTCGCGCCGTTCCCTATTATCCGTTACACCGCCA 2847
1349 LeuThrThrThrGlyAsp 1354
2848 ACTTCGGTAGAATCCCGTTTCAACACGCTGACGGTAAACGGCAAATTGAA 2897

```
      :::::::::::::::      :::|||||||||:::|||:::|||
   1355 SerLysIleAsnAlaThrSerGlyThrLeuThrIleAsnAlaLys..... 1369

   2898 CGGTCAGGGAACATTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCG 2947
                                                      :::
   1370 ..............................................AspA 1371

   2948 ACAAATTGAAGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCGGTC 2997
          ||||||      |||   |||::::::   |||   :::   |||:::
   1371 laLysLeuAspGlyAlaAlaSerGlyAspArgThrValValAsnAlaThr 1387

   2998 AACAATACCGGCAACGAACCTGCAAGCCTCGAACAATTGACGGTAGTGGA 3047
        |||   :::|||:::
   1388 AsnAlaSerGlySer..................................... 1392

   3048 AGGAAAAGACAACAAACCGCTGTCCGAAAACCTTAATTTCACCCTGCAAA 3097
         |||:::   :::         |||:::::::::|||:::|||   :::
   1393 .GlyAsnValThrAlaLysThrSerSerSerValAsnIleThrGlyAspL 1409

   3098 ACGAACACGTCGATGCCGGCGCGTGGCGTTACCAACTCATCCGCAAAGAC 3147
          :::   :::::::::              ::::::|||   ::::::
   1409 euAsnThrIleAsnGly............LeuAsnIleIleSerGluAsn 1421

   3148 GGCGAGTTCCGCCTGCATAATCCGGTCAAAGAACAAGAGCTTTCCGACAA 3197
        |||            :::|||   |||                   ::
   1422 Gly............ArgAsnThrVal.....................Ar 1427

   3198 ACTCGGCAAGGCAGAAGCCAAAAAACAGGCGGAAAAAGACAACGCGCAAA 3247
        :|||         |||         :::      :::   ::::::   |
   1427 gLeuArgGlyLysGluIleAspValLysTyrIleGlnProGlyValAlaS 1444

   3248 GCCTTGACGCGCTGATTGCGGCCGGGCGCGATGCCGTCGAAAAGACAGAA 3297
        ||:::::   :::|||   |||   |||   ::::::||||||::::::
   1444 erValGluGluValIleGluAlaLysArg...ValLeuGluLysValLys 1459

   3298 AGCGTTGCCGAACCGGCCCGGCAGGCAGGCGGGGAAAATGTCGGCATTAT 3347
        ::::::::::::      |||::::::   ::::::   :::|||:::
   1460 AspLeuSerAspGluGluArgGluThrLeuAlaLys...LeuGlyValSe 1475

   3348 GCAGGCGGAGGAAGAGAAAAAACGGGTGCAGGCGGATAAAGACACCGCCT 3397
           :::            :::   |||:::   ::::::      ::::
   1475 rAlaVal............ArgPheValGluProAsnAsnAlaIleThrV 1488

   3398 TGGCGAAACAGCGCGAAGCGGAAACCCGGCCGGCTACC 3435
        ::       |||:::|||       |||:::|||::::::
   1488 alAsnThrGlnAsnGluPheThrThrLysProSerSer 1500

seq_name: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-719-641-10

seq_documentation_block:
; Sequence 10, Application US/08719641
; Patent No. 6218141
;  GENERAL INFORMATION:
;    APPLICANT:  Barenkamp, Stephen J
;    TITLE OF INVENTION:  High Molecular Weight Surface Proteins
;    TITLE OF INVENTION:  of No. 6218141-Typeable Haemophilus
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Shoemaker and Mattare, Ltd.
;      STREET:  2001 Jefferson Davis Hwy., 1203 Crystal Plaza
;      STREET:  Bldg. 1
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202-0286
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/719,641
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/302,832
;      FILING DATE:  16-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US PCT/US93/02166
;      FILING DATE:  16-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9205704.1
;      FILING DATE:  16-MAR-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Berkstresser, Jerry W
;      REGISTRATION NUMBER:  22,651
;      REFERENCE/DOCKET NUMBER:  1038-625
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 415-0810
;      TELEFAX:  (703) 415-0813
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1529 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-719-641-10


alignment_scores:
            Quality:  295.50              Length:   1196
              Ratio:  0.505                 Gaps:     62
 Percent Similarity:  48.913   Percent Identity: 20.819

alignment_block:
 US-09-303-518D-649 x US-08-719-641-10   ..

 Align seg 1/1  to: US-08-719-641-10  from: 1  to: 1529

      289 TCGCGTAACGGCGTGGCGGCATTGGTGGGCGATCAATATATTGTGAGCGT 338
          :::||||||::::::   |||   :::      |||      :::::||
      507 AlaArgAsnAlaThrAspAlaGlnIleThrAlaGlnGlyThrIleThrVa 523

      339 GGCACATAACGGC.........GGCTATAACAACGTTGATTTTGGTGCGG 379
          |      :::                :::||||||||||:::::::::::
      523 lAsnLysAspAspLysGlnPheArgPheAsnAsnValSerIleAsnGlyT 540

      380 AAGGAAGAAATCCCGATCAACATCGTTTTACTTATAAAATTGTGAAACGG 429
            |||::::::            :::|||         |||:::::::::
      540 hrGlyLysGly.........LeuLysPhe.........IleAlaAsnGln 550

      430 AATAATTATAAAGCAGGGACTAAAGGCCATCCTTATGGCGGCGATTATCA 479
          |||||||:::                   |||   :::   |||:::   ::
      551 AsnAsnPheThr..............HisLysPheAspGlyGluLeuAs 562

      480 TATGCCGCGTTTGCATAAATTTGTCACAGATGCAGAACCTGTTGAAATGA 529
          :
      562 n..................................................... 562

      530 CCAGTTATATGGATGGGCGGAAATATATCGATCAAAATAATTACCCTGAC 579
                  ::::::|||          |||:::|||::::::      |||
      563 ........IleSerGlyIleValThrIleAsnGlnThrThrLysLysAsp 576

      580 CGTGTTCGTATTGGGGCAGGCAGGCAATATTGGCGATCTGATGAAGATGA 629
                                   :::||||||:::::::::::::|||::
      577 .....................ValLysTyrTrpAsnAlaSerLysAspSe 586

      630 GCCCAATAACCGCGAAAGT...TCATATCATATTGCAAGTGCGTATTCTT 676
          :      |||   :::|||   :::   :::   ::::::   :::::::
      586 rTyrTrpAsnValSerSerLeuThrLeuAsnThrValGlnLysPheThrP 603

      677 GGCTC............GTTGGTGGCAAT..................... 693
          :::::                |||:::|||
      603 heIleLysPheValAspSerGlySerAsnSerGlnAspLeuArgSerSer 619

      694 ......ACCTTTGCACAA.........AATGGATCAGGTGGTGGCACA.. 726
```

```
620 ArgArgSerPheAlaGlyValHisPheAsnGlyIleGlyGlyLysThrAs 636

727 .GTCAACTTAGGTAGTGAA..............AAAATTAAACATAGC. 759
636 nPheAsnIleGlyAlaAsnAlaLysAlaLeuPheLysLeuLysProAsnA 653

760 ...........CCATATGGTTTTTTACCA..................... 777
653 laAlaThrAspProLysLysGluLeuProIleThrPheAsnAlaAsnIle 669

778 ......ACAGGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATCTA 821
670 ThrAlaThrGlyAsnSer.........AspSerSerValMetPheAspIl 683

822 TGATGCCCAAAAGCAAAAGTGGTTA............ATTAATGGGGTAT 859
683 eHisAlaAsnLeuThrSerArgAlaAlaGlyIleAsnMetAspSerIleA 700

860 TGCAAACGGGC.................AACCCCTATATAGGAAAAAGC 891
700 snIleThrGlyGlyLeuAspPheSerIleThrSerHisAsnArgAsnSer 716

892 AATGGCTTCCAGCTGGTTCGT.......................... 912
717 AsnAlaPheGluIleLysLysAspLeuThrIleAsnAlaThrGlySerAs 733

913 ..................AAAGATTGGTTCTATGATGAAATCTTTGCTG 943
733 nPheSerLeuLysGlnThrLysAspSerPheTyrAsnGluTyrSerLysH 750

944 GAGATACCCATTCAGTATTCTACGAACCACGTCAAAATGGGAAATACTCT 993
750 isAlaIleAsnSerSerHisAsnLeuThrIleLeuGlyGlyAsnValThr 766

994 TTTAACGACGATAATAATGGC.........ACAGGAAAAATCAATGCCAA 1034
767 LeuGlyGlyGluAsnSerSerSerSerIleThrGlyAsnIleAsnIleTh 783

1035 ACATGAACACAAT.......................TCTCTGCCTAATA 1060
783 rAsnLysAlaAsnValThrLeuGlnAlaAspThrSerAsnSerAsnThrG 800

1061 GATTAAAAACACGAACCGTTCAATTGTTTAATGTTTCTTTATCCGAGACA 1110
800 lyLeuLysLysArgThrLeuThrLeuGlyAsnIleSerVal......... 813

1111 GCAAGAGAACCTGTTTATCATGCTGCAGGTGGTGTCAACAGTTATCGACC 1160
814 ......GluGlyAsnLeuSerLeuThrGlyAlaAsnAlaAsnIleValGl 828

1161 CAGACTGAATAATGGAGAAAATATTTCCTTTATTGACGAAGGAAAAGGCG 1210
828 yAsnLeuSerIleAlaGluAspSerThrPheLysGlyGluAlaSerAspA 845

1211 AATTGATACTT.......................ACCAGCAACATCAAT 1236
845 snLeuAsnIleThrGlyThrPheThrAsnAsnGlyThrAlaAsnIleAsn 861

1237 CAAGGTGCTGGAGGATTATATTTCCAAGGAGATTTTACGGTCTCGCCTGA 1286
862 IleLysGlnGlyValValLysLeuGlnGlyAspIle.............. 873

1287 AAATAACGAAACTTGGCAAGGCGCGGGCGTTCATATCAGTGAAGACAGTA 1336
874 .AsnAsnLys............GlyGlyLeuAsnIleThrThrAsnAlaS 886

1337 CCGTTACTTGGAAA......GTAAACGGC...GTGGCAAACGACCGCCTG 1377
886 erGlyThrGlnLysThrIleIleAsnGlyAsnIleThrAsnGlu...... 900

1378 TCCAAAATCGGCAAAGGCACGCTG.........CACGTTCAAGCCAAAGG 1418
```

```
901 ...........LysGlyAspLeuAsnIleLysAsnIleLysAlaAspAl 913

1419 GGAAAACCAA......GGCTCGATCAGCGTGGGCGACGGTACAGTCATTT 1462
913 aGluIleGlnIleGlyGlyAsnIleSerGlnLysGluGlyAsnLeuThrI 930

1463 TG.............................................. 1464
930 leSerSerAspLysValAsnIleThrAsnGlnIleThrIleLysAlaGly 946

1465 .....................GATCAGCAGGCAGACGATAAAGGCAAAAA 1493
947 ValGluGlyGlyArgSerAspSerSerGluAlaGluAsnAlaAsnLeuTh 963

1494 ACAAGCCTTTAGTGAAATCGGCTTG..............GTCAGCGGCA 1528
963 rIleGlnThrLysGluLeuLysLeuAlaGlyAspLeuAsnIleSerGlyP 980

1529 GGGGTACGGTGCAACTGAATGCCGATAAT..................... 1557
980 heAsnLysAlaGluIleThrAlaLysAsnGlySerAspLeuThrIleGly 996

1558 ..............CAGTTCAACCCCGACAAACTCTATTTCGGC...TT 1589
997 AsnAlaSerGlyGlyAsnAlaAspAlaLysLysValThrPheAspLysVa 1013

1590 TCGCGGCGGACGTTTGGATTTAAACGGGCATTCGCTTTCGTTCCACCGTA 1639
1013 lLysAspSerLysIleSerThrAspGlyHisAsnValThrLeuAsnSerG 1030

1640 TTCAAAATACCGATGAAGGGGCGATGATTGTCAACCACAATCAAGACAAA 1689
1030 luValLysThrSerAsnGlySerSer......AsnAlaGlyAsnAspAsn 1044

1690 GAATCCACCGTTACCATTACAGGCAATAAAGATATTGCTACAACCGGCAA 1739
1045 SerThrGlyLeuThrIleSerAla...LysAspValThrVal........ 1057

1740 TAACAACAGCTTGGATAGCAAAAAAGAAATTGCCTACAACGGTTGGTTTG 1789
1058 .AsnAsnAsnValThrSerHisLysThrIleAsnIleSerAlaAlaAlaG 1074

1790 GCGAGAAAGATACGACCAAAACGAACGGGCGGCTCAACCTTGTTTACCAG 1839
1074 lyAsnVal...ThrThrLysGluGlyThrThrIleAsn............ 1085

1840 CCCGCCGCAGAAGACCGCACCCTGCTGCTTTCCGGCGGAACAAATTTAAA 1889
1086 ...AlaThrThrGlySerValGluValThrAlaGlnAsnGlyThrIleLy 1101

1890 CGGCAACATCACGCAAACAAACGGCAAACTGTTTTTCAGCGGCAGACCAA 1939
1101 sGlyAsnIleThrSerGlnAsn.....................ValT 1110

1940 CACCGCACGCCTACAATCATTTAAACGACCATTGGTCGCAAAAAGAGGGC 1989
1110 hrValThrAlaThrGluAsnLeu..................... 1117

1990 ATTCCTCGCGGGGAAATCGTGTGGGACAACGACTGGATCAACCGCACATT 2039
1118 ..............ValThrThrGluAsnAlaValIleAsnAlaThrSe 1129

2040 TAAAGCGGAAAACTTCCAAATTAAAGGCGGACAGGCGGTGGTTTCCCGCA 2089
1129 rGlyThrValAsnIleSerThrLysThrGlyAsp......IleLysGlyG 1144

2090 ATGTTGCCAAAGTGAAAGGCGATTGGCATTTGAGCAATCACGCCCAAGCA 2139
1144 lyIleGluSerThrSerGlyAsnValAsnIleThrAlaSerGlyAsnThr 1160

2140 GTTTTTGGTGTCGCACCGCATCAAAGCCACACAATCTGTACACGTTCGGA 2189
1161 Leu...LysValSerAsnIleThrGlyGlnAspValThrValThrAlaAs 1176
```

```
2190 CTGGACGGGTCTGACAAATTGTGTCGAAAAAACCATTACCGACGATAAAG 2239
1176 pAlaGlyAlaLeuThrThrThrAlaGlySerThrIleSerAlaThrThrG 1193

2240 TGATTGCTTCATTG...ACTAAGACC...GACATCAGCGGCAATGTCGAT 2283
1193 lyAsnAlaAsnIleThrThrLysThrGlyAspIleAsnGlyLysValGlu 1209

2284 CTTGCCGAT...CACGCTCATTTAAATCTCACAGGGCTTGCCACACTCAA 2330
1210 SerSerSerGlySerValThrLeuValAlaThrGlyAlaThrLeuAlaVa 1226

2331 CGGCAATCTTAGTGCAAATGGCGATACA.................CGTT 2362
1226 lGlyAsnIleSerGlyAsnThrValThrIleThrAlaAspSerGlyLysL 1243

2363 ATACAGTCAGCCACAACGCCACCCAAAACGGCAACCTTAGCCTCGTGGGC 2412
1243 euThrSerThrValGlySerThrIleAsnGlyThrAsnSerValThrThr 1259

2413 AATGCCCAAGCAACATTTAATCAAGCCACATTAAACGGCAACACATCGGC 2462
1260 SerSerGlnSerGlyAspIleGluGlyThrIleSerGlyAsnThr..... 1274

2463 TTCGGGCAATGCTTCATTTAATCTAAGCGACCACGCCGTACAAAACGGCA 2512
1275 ...............ValAsnValThr.........AlaSerThrGlyA 1283

2513 GTCTGACGCTTTCCGGCAACGCTAAGGCAAACGTAAGCCATTCCGCACTC 2562
1283 spLeuThrIleGlyAsnSerAlaLysValGluAlaLysAsnGlyAla... 1298

2563 AACGGTAATGTCTCCCTAGCCGATAAGGCAGTATTCCATTTTGAAAGCAG 2612
1299 .........................AlaThrLeuThrAlaGluSerGl 1306

2613 CCGCTTTACCGGACAA..............ATCAGCGGCGGCAAGGATA 2647
1306 yLysLeuThrThrGlnThrGlySerSerIleThrSerSerAsnGlyGlnT 1323

2648 CGGCATTACACTTAAAAGACAGCGAATGGACGCTGCCGTCAGGCACGGAA 2697
1323 hrThrLeuThrAlaLysAspSer.....................SerIle 1332

2698 TTAGGCAATTTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTA 2747
1333 AlaGlyAsnIleAsnAla......AlaAsnValThrLeuAsnThr..... 1345

2748 TCGCCACGATGCGGCAGGGGCGCAAACCGGCAGTGCGACAGATGCGCCGC 2797
1346 .........................ThrGlyThr............... 1348

2798 GCCGCCGTTCGCGCCGTTCGCGCCGTTCCCTATTATCCGTTACACCGCCA 2847
1349 ...............................LeuThrThrThrGlyAsp 1354

2848 ACTTCGGTAGAATCCCGTTTCAACACGCTGACGGTAAACGGCAAATTGAA 2897
1355 SerLysIleAsnAlaThrSerGlyThrLeuThrIleAsnAlaLys..... 1369

2898 CGGTCAGGGAACATTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCG 2947
1370 .............................................AspA 1371

2948 ACAAATTGAAGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCGGTC 2997
1371 laLysLeuAspGlyAlaAlaSerGlyAspArgThrValValAsnAlaThr 1387

2998 AACAATACCGGCAACGAACCTGCAAGCCTCGAACAATTGACGGTAGTGGA 3047
1388 AsnAlaSerGlySer................................ 1392

3048 AGGAAAAGACAACAAACCGCTGTCCGAAAACCTTAATTTCACCCTGCAAA 3097
1393 .GlyAsnValThrAlaLysThrSerSerSerValAsnIleThrGlyAspL 1409

3098 ACGAACACGTCGATGCCGGCGCGTGGCGTTACCAACTCATCCGCAAAGAC 3147
1409 euAsnThrIleAsnGly............LeuAsnIleIleSerGluAsn 1421

3148 GGCGAGTTCCGCCTGCATAATCCGGTCAAAGAACAAGAGCTTTCCGACAA 3197
1422 Gly............ArgAsnThrVal.....................Ar 1427

3198 ACTCGGCAAGGCAGAAGCCAAAAAACAGGCGGAAAAAGACAACGCGCAAA 3247
1427 gLeuArgGlyLysGluIleAspValLysTyrIleGlnProGlyValAlaS 1444

3248 GCCTTGACGCGCTGATTGCGGCCGGGCGCGATGCCGTCGAAAAGACAGAA 3297
1444 erValGluGluValIleGluAlaLysArg...ValLeuGluLysValLys 1459

3298 AGCGTTGCCGAACCGGCCCGGCAGGCAGGCGGGGAAAATGTCGGCATTAT 3347
1460 AspLeuSerAspGluGluArgGluThrLeuAlaLys...LeuGlyValSe 1475

3348 GCAGGCGGAGGAAGAGAAAAAACGGGTGCAGGCGGATAAAGACACCGCCT 3397
1475 rAlaVal............ArgPheValGluProAsnAsnAlaIleThrV 1488

3398 TGGCGAAACAGCGCGAAGCGGAAACCCGGCCGGCTACC 3435
1488 alAsnThrGlnAsnGluPheThrThrLysProSerSer 1500
```

```
seq_name: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-460-269C-6

seq_documentation_block:
; Sequence 6, Application US/08460269C
; Patent No. 6197548
;   GENERAL INFORMATION:
;        APPLICANT: CLARE, JEFFREY J.
;                   ROMANOS, MICHAEL A.
;        TITLE OF INVENTION: EXPRESSION OF HETEROLOGOUS PROTEIN IN
;                            YEAST
;        NUMBER OF SEQUENCES: 17
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Millen, White, Zelano & Branigan, P.C.
;             STREET: 2200 Clarendon Blvd., Suite 1400
;             CITY: ARLINGTON
;             STATE: VA
;             COUNTRY: USA
;             ZIP: 22201
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/08/460,269C
;             FILING DATE: 02-Jun-1995
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Lebovitz, Richard M.
;             REGISTRATION NUMBER: 37,067
;             REFERENCE/DOCKET NUMBER: Popov-2
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (703) 243-6333
;             TELEFAX: (703) 243-6410
;   INFORMATION FOR SEQ ID NO: 6:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 922 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-08-460-269C-6
```

```
alignment_scores:
            Quality:  295.00              Length:   1083
              Ratio:  0.602                Gaps:      52
 Percent Similarity:  45.245   Percent Identity: 22.253

alignment_block:
 US-09-303-518D-649 x US-08-460-269C-6   ..

 Align seg 1/1  to: US-08-460-269C-6  from: 1  to: 922

    1417 GGGGAAAACCAAGGCTCGATCAGCGTGGGCGACGGTACAGTCATTTTGGA 1466
         ||||||:::|||   :::|||   :::
      45 GlyGluArgGlnHisGlyIleHisIle..................... 53

    1467 TCAGCAGGCAGACGATAAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCT 1516
          :::|||:::|||      |||   :::   |||   :::
      54 .LysGlnSerAspGlyAlaGlyValArgThrAlaThrGlyThrThrIleL 70

    1517 TGGTCAGCGGCAGGGGTACGGTGCAACTGAATGCCGATAATCAGTTCAAC 1566
           ||||||||||||   :::   :::      :::|||
      70 ysValSerGlyArgGlnAlaGlnGlyValLeuLeuGluAsn......... 83

    1567 CCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAACGG 1616
                    :::   |||:::::::|||   :::      :::||
      84 ...ProAlaAlaGluLeuArgPheGlnAsnGlySerValThrSerSerGl 99

    1617 GCATTCGCTTTCGTTCCACCGTATTCAAAATACCGATGAAGGGGCGATGA 1666
         |                        |||      |||||||||:::
      99 y.......................GlnLeuPheAspGluGlyVal.... 106

    1667 TTGTCAACCACAATCAAGACAAAGAATCCACCGTTACCATTACAGGCAAT 1716
                      :::         :::|||||||||:::   ::::::
     107 ..............ArgArgPheLeuGlyThrValThrValLysAlaGly 118

    1717 AAA.........GATATTGCTACAACCGGCAATAACAACAGCTTGGATAG 1757
         |||        |||   ||||||   :::|||
     119 LysLeuValAlaAspHisAlaThrLeuAlaAsn................. 129

    1758 CAAAAAAGAAATTGCCTACAACGGTTGGTTTGGCGAGAAAGATACGACCA 1807
                                           :::||||||
     130 ..................................ValSerAspThrArgA 135

    1808 AAACGAACGGGCGGCTCAACCTTGTT............TACCAGCCCGCC 1845
               :::|||            |||              |||   :::
     135 spAspAspGlyIleAlaLeuTyrValAlaGlyGluGlnAlaGlnAlaSer 151

    1846 GCAGAAGACCGCACCCTGCTGCTTTCCGGCGGAACAAATTTA......AA 1889
               |||   ||||||      :::||||||:::::::::      ::
     152 IleAlaAspSerThrLeuGlnGlyAlaGlyGlyValArgValGluArgGl 168

    1890 CGGCAACATCACGCAAACAAACGGCAAACTGTTTTTCAGCGGC....... 1932
         ::::|||:::|||      ::::::   :::      :::|||
     168 yAlaAsnValThrValGlnArgSerThrIleValAspGlyGlyLeuHisI 185

    1933 ..........AGACCAACACCGCACGCCTACAATCATTTAAACGACCAT 1971
                    :::|||
     185 leGlyThrLeuGlnProLeu............................. 191

    1972 TGGTCGCAAAAAGAGGGCATTCCTCGCGGGGAAATCGTGTGGGACAACGA 2021
               |||   |||   :::|||   ::::::::::|||      :::
     192 ......GlnProGluAspLeuProProSerArgValValLeuGlyAspTh 206

    2022 CTGGATCAACCGC..............................ACAT 2038
               ::::::                              :::|
     206 rSerValThrAlaValProAlaSerGlyAlaProAlaAlaValPheValP 223

    2039 TTAAAGCGGAAAACTTCCAAATTAAAGGCGGACAGGCGGTGGTTTCCCGC 2088
         ||   |||::::::::::   :::   ||||||:::   :::   :::|||
     223 heGlyAlaAsnGluLeuThrValAspGlyGlyHisIleThrGlyGlyArg 239

    2089 AATGTTGCCAAAGTGAAAGGCGATTGGCATTTGAGCAATCACGCCCAAGC 2138
```

```
                :::::                                  :::   ||
     240 AlaAlaGly.................................ValAlaAl 245

    2139 AGTTTTTGGTGTCGCACCGCATCAAAGCCACACAATCTGTACACGTTCGG 2188
         |:::   |||:::      |||   :::::::::         |||:::|
     245 aMetAspGlyAlaIleValHisLeuGlnArgAlaThrIleArgArgGlyA 262

    2189 AC.............................................. 2190
         ||
     262 spAlaProAlaGlyGlyAlaValProGlyGlyAlaValProGlyGlyAla 278

    2191 ..............................TGGACGGGTCTGACAAA 2207
                                        |||   |||:::
     279 ValProGlyGlyPheGlyProLeuLeuAspGlyTrpTyrGlyVal..... 293

    2208 TTGTGTCGAAAAAACCATTACCGACGATAAAGTGATTGCTTCATTGACTA 2257

     293 ............................................... 293

    2258 AGACCGACATCAGC...GGCAATGTCGATCTTGCCGATCACGCTCATTTA 2304
               |||:::|||   ::::::|||||||||||||:::
     294 .....AspValSerAspSerThrValAspLeuAlaGln............ 304

    2305 AATCTCACAGGGCTTGCCACACTCAACGGCAATCTTAGTGCAAATGGCGA 2354
         :::::::::            |||::::::   :::   |||:::
     305 SerIleValGluAlaProGlnLeuGlyAlaAlaIleArgAlaGlyArgGl 321

    2355 TACACGTTATACAGTCAGCCACAAC......GCCACCCAAAACGGCAACC 2398
          :::|||   ||||||||   :::      ::::::   :::|||||||:
     321 yAlaArgValThrValSerGlyGlySerLeuSerAlaProHisGlyAsnV 338

    2399 TTAGCCTCGTGGGCAATGCCCAAGCAACATTTAAT............... 2433
         ::      :::|||::::::         |||
     338 alIleGluThrGlyGlyGlyAlaArgArgPheProProProAlaSerPro 354

    2434 ...CAAGCCACATTAAACGGCAACACATCGGCTTCGGGCAATGCT...TC 2477
            :::   ||||||:::::::::::   |||:::|||:::|||
     355 LeuSerIleThrLeuGlnAlaGlyAlaArgAlaGlnGlyArgAlaLeuLe 371

    2478 ATTTAATCTAAGCGACCACGCCGTACAAAACGGCAGTCTGACGCTTTCCG 2527
          :::::::::      :::   |||:::          |||||||||:::|
     371 uTyrArgValLeuProGluProValLys.........LeuThrLeuAlaG 385

    2528 GCAACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACGGTAATGTCTCC 2577
         ||:::|||::::::                        :::|||::::::
     385 lyGlyAlaGlnGly.....................GlnGlyAspIleVal 394

    2578 CTAGCCGATAAGGCAGTATTCCATTTTGAAAGCAGCCGCTTTACCGGACA 2627
            :::                     |||         :::   |||
     395 AlaThr.....................GluLeuProProIleProGlyAl 404

    2628 AATCAGCGGCGGCAAGGATACGGCATTACACTTAAAAGACAGCGAATGGA 2677
              ||||||      |||:::|||||||      :::::::::::||||
     404 aSerSerGlyProLeuAspValAlaLeu...AlaSerGlnAlaArgTrpT 420

    2678 CGCTGCCGTCAGGCACGGAATTAGGCAATTTAAACCTTGACAACGCCACC 2727
         ||      :::         :::   :::|||::::::|||||||||||
     420 hr...GlyAlaThrArgAlaValAspSerLeuSerIleAspAsnAlaThr 435

    2728 ATTACACTCAATTCCGCCTATCGCCACGATGCGGCAGGGGCGCAAACCGG 2777
            ::::::
     436 TrpValMet.......................................... 438

    2778 CAGTGCGACAGATGCGCCGCGCCGCCGTTCGCGCCGTTCGCGCCGTTCCC 2827
                  ||||||                ::::|||:::   :::
     439 .......ThrAsp...................AsnSerAsnValGlyA 446

    2828 TATTATCCGTTACACCGCCAACTTCGGTAGAA.................. 2859
           |||   ::::::         ||||||:::
     446 laLeuArgLeuAlaSerAspGlySerValAspPheGlnGlnProAlaGlu 462

    2860 ...TCCCGTTTCAACACGCTGACGGTAAACGGCAAATTGAACGGTCAGGG 2906
            :::|||||||::::::|||   |||:::      |||   |||:::||
```

```
 463 AlaGlyArgPheLysValLeuMetValAsp...ThrLeuAlaGlySerGl 478

2907 AACATTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGACAAATTGA 2956
     |   ||||||:::                              ||||||||||||
 478 yLeuPheArgMetAsnValPheAlaAspLeuGlyLeuSerAspLysLeuV 495

2957 AGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCGGTCAACAATACC 3006
       :::   ::::::::::::|||   :::   |||   |||:::|||:::
 495 alValMetArgAspAlaSerGlyGlnHisArgLeuTrpValArgAsnSer 511

3007 GGCAACGAACCTGCAAGCCTCGAACAATTGACGGTAGTGGAAGGAAAAGA 3056
     |||:::||||||||||||   :::   :::   :::|||:::
 512 GlySerGluProAlaSerGlyAsnThrMetLeuLeuValGln........ 525

3057 CAACAAACCGCTGTCCGAAAACCTTAATTTCACCCTGCAAAACGAA.... 3102
          |||   ::::::      :::|||||||||   |||:::
 526 ....ThrProArgGlySerAlaAlaThrPheThrLeuAlaAsnLysAspG 541

3103 ..CACGTCGATGCCGGCGCGTGGCGTTACCAACTCATCCGCAAAGACGGC 3150
       ||||||   |||:::::|||||||:::|||
 541 lyLysValAspIleGlyThrTyrArgTyrArgLeu.............. 552

3151 GAGTTCCGCCTGCATAATCCGGTCAAAGAACAAGAGCTTTCCGACAAACT 3200

 552 .................................................. 552

3201 CGGCAAGGCAGAAGCCAAAAAACAGGCGGAAAAAGACAACGCGCAAAGCC 3250
                          |||   :::   |||:::|||
 553 ........................AlaAlaAsnGlyAsnGlyGlnTrp. 560

3251 TTGACGCGCTGATTGCGGCCGGGCGCGATGCCGTCGAAAAGACAGAAAGC 3300
       :::|||:::::|||   |||   |||:::         |||
 561 .....SerLeuValGlyAla.LysAlaProProAla......ProLysPr 573

3301 GTTGCCGAACCGGCCCGGCAGGCAGGCGGGGAAAATGTCGGCATTATGCA 3350
       |||   :::||||||   |||   |||   :::
 573 oAlaPro...GlnProGlyProGlnProGlyProGln.............. 584

3351 GGCGGAGGAAGAGAAAAAACGGGTGCAGGCGGATAAAGACACCGCCTTGG 3400
                                          ||||||
 585 ........................................ProPro... 586

3401 CGAAACAGCGCGAAGCGGAAACCCGGCCGGCTACCACCGCCTTCCCCCGC 3450
                                    ||||||    ||||||:
 587 ..............................GlnProProGln.ProProG 593

3451 GCCCGCCGCGCCCGCCGGGATTTGCCGCAACTGCAACCCCAA........ 3492
     ::||||||:::|||||| :::   ||||||   |||||||:::
 593 lnProProGlnProPro.GlnProProGlnArgGlnProGluAlaProAl 609

3493 .CCGCAGCCCCAACCGCAGCGCGACCTGATCAGCCGTTATGCCAATAGCG 3541
      |||||||||         |||:::|||   |||      ||||||::::
 609 aProGlnProProAlaGlyArgGluLeu...SerAlaAlaAlaAsnAlaA 625

3542 GTTTGAGTGAATTTTCCGCCACGCTCAACAGCGTT......TTCGCCGTA 3585
     ::::::::      ::::::   |||   |||:::      :::|||
 625 laValAsnThrGlyGlyValGlyLeuAlaSerThrLeuTrpTyrAlaGlu 641

3586 CAGGACGAATTAGACCGCGTATTTGCCGAAGACCGCCGCAAC........ 3627
     ::::::   |||:::::::   :::::::|||   |||   |||
 642 SerAsnAlaLeuSerLysArgLeuGlyGluLeuArgLeuAsnProAspAl 658

3628 ....GCCGTTTGGACAAGCGGCATCCGGGACACCAAACACTACCGTTCGC 3673
         ::::::|||      |||:::   :::   ::::::      ::::
 658 aGlyGlyAlaTrpGlyArgGlyPheAlaGlnArgGlnGlnLeuAspAsnA 675

3674 AAGATTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCAAATCGGTATG 3723
     ::      |||   :::   |||::::::       :::::::|||
 675 rgAlaGlyArgArgPheAspGlnLysValAlaGlyPheGluLeuGlyAla 691

3724 CAGAAAAACCTC............GGCAGCGGGCGCGTCGGCATCCTGTT 3761
     :::      :::            |||      ::::::|||   |||
 692 AspHisAlaValAlaValAlaGlyGlyArgTrpHisLeuGlyGlyLeuAl 708

3762 TTCGCACAACCGGACCGAAAACACCTTC...GACGACGGCATCGGCAACT 3808
      :::::::::|||   ::::::   |||      ||||||   |||::::
 708 aGlyTyrThrArgGlyAspArgGlyPheThrGlyAspGlyGlyGlyHisT 725

3809 CGGCACGGCTTGCCCACGGCGCCGTTTTCGGGCAATACGGCATCGACAGG 3858
     ::      :::      |||:::                :::   :::
 725 hrAspSerValHisValGlyGlyTyrAlaThrTyrIleAlaAsnSerGly 741

3859 TTCTACATCGGCATCAGCGCGGGCGCGGGTTTTAGCAGCGGCAGCCTT.. 3906
     ||||||:::      :::      |||:::      ::::::::::::
 742 PheTyrLeuAspAlaThrLeuArgAlaSerArgLeuGluAsnAspPheLy 758

3907 ..........TCAGACGGC......ATCGGAGGCAAAATCCGCCGCCGCG 3940
                 |||||||||      :::   ||||||   |||   :::
 758 sValAlaGlySerAspGlyTyrAlaValLysGlyLysTyrArgThrHisG 775

3941 TGCTGCATTACGGCATTCAGGCACGATACCGC...GCCGGTTTCGGCGGA 3987
       :::      ::::::::::|||      |||   |||         |||
 775 lyValGlyValSerLeuGluAlaGlyArgArgPheAlaHisAlaAspGly 791

3988 TTCGGCATCGAACCGCACATCGGCGCAACGCGCTATTTCGTCCAAAAAGC 4037
     :::   :::|||||||:::       :::   :::   |||       ::
 792 TrpPheLeuGluProGlnAlaGluLeuAlaValPheArgValGlyGlyGl 808

4038 GGATTACCGCTACGAAAAC...GTCAATATCGCCACCCCCGGCCTTGCAT 4084
     :   ||||||      |||   :::::::::         |||   :::
 808 yAlaTyrArgAlaAlaAsnGlyLeuArgValArgAspGluGlyGlySerS 825

4085 TCAACCGCTACCGCGCGGGCATTAAGGCAGATTATTCATTCAAACCGGCG 4134
                  |||   |||:::::::::         ::::::   |||
 825 erValLeuGlyArgLeuGlyLeuGluValGlyLysArgIleGluLeuAla 841

4135 CAACACATTTCCATCACGCCTTATTTGAGCCTGTCCTATACCGATGCCGC 4184
               ::::::   ||||||::::::   |||      :::
 842 GlyGlyArgGlnValGlnProTyrIleLysAlaSerValLeuGlnGluPh 858

4185 TTCGGGCAAAGTCCGAACACGCGTCAATACCGCCGTATTGGCTCAGGATT 4234
      :::|||         :::|||:::|||      :::         ::::
 858 eAspGlyAlaGlyThrValArgThrAsnGlyIleAlaHisArgThrGluL 875

4235 TCGGCAAAACCCGCAGTGCGGAATGGGGCGTAAACGCCGAAATCAAAGGT 4284
     ::      ||||||   ||||||   |||
 875 euArgGlyThrArg...AlaGluLeuGly..................... 883

4285 TTCACGCTGTCCCTCCACGCTGCCGCCGCCAAAGGCCCGCAACTGGAAGC 4334
            |||::::::   ||||||   :::::::|||   :::|||   ||
 884 ......LeuGlyMetAlaAlaAlaLeuGlyArgGlyHisSerLeuTyrAl 898

4335 G......CAACACAGCGCCGGGCATCAAATTAGGCTACCGCTGG 4371
     |       :::::::|||   |||   ||||||:::      |||
 898 aSerTyrGluTyrSerLysGlyProLysLeuAlaMetProTrp 912

seq_name: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-617-697-10

seq_documentation_block:
; Sequence 10, Application US/08617697
; Patent No. 5977336
;  GENERAL INFORMATION:
;    APPLICANT:  Barenkamp, Stephen J
;    TITLE OF INVENTION:  High Molecular Weight Surface Proteins
;    TITLE OF INVENTION:  of No. 5977336-Typeable Haemophilus
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Shoemaker and Mattare, Ltd.
;      STREET:  2001 Jefferson Davis Hwy., 1203 Crystal Plaza
;      STREET:  Bldg. 1
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202-0286
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
```

```
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/617,697
;      FILING DATE:  01-APR-1996
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/302,832
;      FILING DATE:  05-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US PCT/US93/02166
;      FILING DATE:  16-MAR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Berkstresser, Jerry W
;      REGISTRATION NUMBER:  22,651
;      REFERENCE/DOCKET NUMBER:  1038-557
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 415-0810
;      TELEFAX:  (703) 415-0813
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1600 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-617-697-10


alignment_scores:
            Quality:  292.50              Length:   1189
              Ratio:  0.497                 Gaps:     61
 Percent Similarity:  49.537   Percent Identity: 20.606

alignment_block:
 US-09-303-518D-649 x US-08-617-697-10   ..

 Align seg 1/1  to: US-08-617-697-10  from: 1  to: 1600

     289 TCGCGTAACGGCGTGGCGGCATTGGTGGGCGATCAATATATTGTGAGCGT 338
         :::||||||:::::   |||   :::      |||      :::::||
     579 AlaArgAsnAlaThrAspAlaGlnIleThrAlaGlnGlyThrIleThrVa 595

     339 GGCACATAACGGC.........GGCTATAACAACGTTGATTTTGGTGCGG 379
         |      :::                :::||||||||||:::::::::::
     595 lAsnLysAspAspLysGlnPheArgPheAsnAsnValSerLeuAsnGlyT 612

     380 AAGGAAGAAATCCCGATCAACATCGTTTTACTTATAAAATTGTGAAACGG 429
           |||::::::            :::|||         |||:::::::::
     612 hrGlyLysGly.........LeuLysPhe.........IleAlaAsnGln 622

     430 AATAATTATAAAGCAGGGACTAAAGGCCATCCTTATGGCGGCGATTATCA 479
         ||||||:::                 |||   :::   |||:::   ::
     623 AsnAsnPheThr..............HisLysPheAspGlyGluIleAs 634

     480 TATGCCGCGTTTGCATAAATTTGTCACAGATGCAGAACCTGTTGAAATGA 529
         :
     634 n..................................................  634

     530 CCAGTTATATGGATGGGCGGAAATATATCGATCAAAATAATTACCCTGAC 579
                 ::::::|||         |||:::|||::::::      |||
     635 ........IleSerGlyIleValThrIleAsnGlnThrThrLysLysAsp 648

     580 CGTGTTCGTATTGGGGCAGGCAGGCAATATTGGCGATCTGATGAAGATGA 629
                               :::||||||:::::::::::::|||::
     649 .....................ValLysTyrTrpAsnAlaSerLysAspSe 658

     630 GCCCAATAACCGCGAAAGT...TCATATCATATTGCAAGTGCGTATTCTT 676
         :      |||   :::|||   :::   :::   ::::::   :::::::
     658 rTyrTrpAsnValSerSerLeuThrLeuAsnThrValGlnLysPheThrP 675

     677 GGCTC............GTTGGTGGCAAT..................... 693
         :::::               |||:::|||
```

```
     675 heIleLysPheValAspSerGlySerAsnGlyGlnAspLeuArgSerSer 691

     694 ......ACCTTTGCACAA.........AATGGATCAGGTGGTGGCACA.. 726
               :::||||||          ||||||   ||||||   |||
     692 ArgArgSerPheAlaGlyValHisPheAsnGlyIleGlyGlyLysThrAs 708

     727 .GTCAACTTAGGTAGTGAA..............AAAATTAAACATAGC. 759
          |||:::|||::::::                |||:::|||   :::
     708 nPheAsnIleGlyAlaAsnAlaLysAlaLeuPheLysLeuLysProAsnA 725

     760 ...........CCATATGGTTTTTTACCA..................... 777
                 |||         ||||||
     725 laAlaThrAspProLysLysGluLeuProIleThrPheAsnAlaAsnIle 741

     778 ......ACAGGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATCTA 821
               ||||||:::|||         ::::::|||   ||||||
     742 ThrAlaThrGlyAsnSer.........AspSerSerValMetPheAspIl 755

     822 TGATGCCCAAAAGCAAAAGTGGTTA............ATTAATGGGGTAT 859
          |||:::      :::                    ::::::::::::
     755 eHisAlaAsnLeuThrSerArgAlaAlaGlyIleAsnMetAspSerIleA 772

     860 TGCAAACGGGC..................AACCCCTATATAGGAAAAAGC 891
          ||||||                    :::   :::      :::|||
     772 snIleThrGlyGlyLeuAspPheSerIleThrSerHisAsnArgAsnSer 788

     892 AATGGCTTCCAGCTGGTTCGT........................... 912
         |||:::|||::::::   :::
     789 AsnAlaPheGluIleLysLysAspLeuThrIleAsnAlaThrGlySerAs 805

     913 ...................AAAGATTGGTTCTATGATGAAATCTTTGCTG 943
                             ||||||   ||||||:::|||
     805 nPheSerLeuLysGlnThrLysAspSerPheTyrAsnGluTyrSerLysH 822

     944 GAGATACCCATTCAGTATTCTACGAACCACGTCAAAATGGGAAATACTCT 993
               :::|||                      :::|||:::   :::
     822 isAlaIleAsnSerSerHisAsnLeuThrIleLeuGlyGlyAsnValThr 838

     994 TTTAACGACGATAATAATGGC.........ACAGGAAAAATCAATGCCAA 1034
         ::::::   :::|||:::::          ||||||:::||||||
     839 LeuGlyGlyGluAsnSerSerSerIleThrGlyAsnIleAsnIleTh 855

    1035 ACATGAACACAAT.......................TCTCTGCCTAATA 1060
          ::::::   |||                      :::      :::
     855 rAsnLysAlaAsnValThrLeuGlnAlaAspThrSerAsnSerAsnThrG 872

    1061 GATTAAAAACACGAACCGTTCAATTGTTTAATGTTTCTTTATCCGAGACA 1110
           ||||||   ||||||:::   |||   |||:::|||:::
     872 lyLeuLysLysArgThrLeuThrLeuGlyAsnIleSerVal......... 885

    1111 GCAAGAGAACCTGTTTATCATGCTGCAGGTGGTGTCAACAGTTATCGACC 1160
          |||                :::|||:::      :::
     886 ......GluGlyAsnLeuSerLeuThrGlyAlaAsnAlaAsnIleValGl 900

    1161 CAGACTGAATAATGGAGAAAATATTTCCTTTATTGACGAAGGAAAAGGCG 1210
          :::|||:::   :::|||:::   :::|||      |||::::::   :
     900 yAsnLeuSerIleAlaGluAspSerThrPheLysGlyGluAlaSerAspA 917

    1211 AATTGATACTTACCAGCAACATC...AATCAAGGTGCTGGAGGATTATAT 1257
         ::|||   :::|||:::::::::   |||:::|||:::::::::::::
     917 snLeuAsnIleThrGlyThrPheThrAsnAsnGlyThrAlaAsnIleAsn 933

    1258 TTCCAAGGAGATTTTACGGTCTCGCCTGAAAATAACGAAACTTGGCAAGG 1307
         :::::::|||         ::::::      ||||||:::
     934 IleLysGlyValValLysLeuGlyAspIleAsnAsnLys........... 946

    1308 CGCGGGCGTTCATATCAGTGAAGACAGTACCGTTACTTGGAAA......G 1351
          :::|||::::::|||:::   :::::::::   |||   |||      :
     947 .GlyGlyLeuAsnIleThrThrAsnAlaSerGlyThrGlnLysThrIleI 963

    1352 TAAACGGC...GTGGCAAACGACCGCCTGTCCAAAATCGGCAAAGGCACG 1398
         ::||||||   ::::::|||:::                    ||||||
     963 leAsnGlyAsnIleThrAsnGlu.................LysGlyAsp 973
```



```
                                        |||:::   :::
    1463 ..............................GlyAsnValThrAlaLysTh 1469

    3069 GTCCGAAAACCTTAATTTCACCCTGCAAAACGAACACGTCGATGCCGGCG 3118
          |||::::::::::|||:::|||   :::   :::   :::::::::
    1469 rSerSerSerValAsnIleThrGlyAspLeuAsnThrIleAsnGly.... 1484

    3119 CGTGGCGTTACCAACTCATCCGCAAAGACGGCGAGTTCCGCCTGCATAAT 3168
                  :::::::|||   :::::::|||            :::|||
    1485 ........LeuAsnIleIleSerGluAsnGly............ArgAsn 1494

    3169 CCGGTCAAAGAACAAGAGCTTTCCGACAAACTCGGCAAGGCAGAAGCCAA 3218
           |||                          :::|||         |||
    1495 ThrVal.....................ArgLeuArgGlyLysGluIleAs 1504

    3219 AAAACAGGCGGAAAAAGACAACGCGCAAAGCCTTGACGCGCTGATTGCGG 3268
          :::      :::   ::::::   |||::::::   :::|||   |
    1504 pValLysTyrIleGlnProGlyValAlaSerValGluGluValIleGluA 1521

    3269 CCGGGCGCGATGCCGTCGAAAAGACAGAAAGCGTTGCCGAACCGGCCCGG 3318
          ||   |||   ::::::||||||:::::::::::::::::::      |||
    1521 laLysArg...ValLeuGluLysValLysAspLeuSerAspGluGluArg 1536

    3319 CAGGCAGGCGGGGAAAATGTCGGCATTATGCAGGCGGAGGAAGAGAAAAA 3368
          ::::::   ::::::   :::|||:::      :::             ::
    1537 GluThrLeuAlaLys...LeuGlyValSerAlaVal............Ar 1548

    3369 ACGGGTGCAGGCGGATAAAGACACCGCCTTGGCGAAACAGCGCGAAGCGG 3418
          :   |||:::   ::::::      ::::::      |||:::|||
    1548 gPheValGluProAsnAsnAlaIleThrValAsnThrGlnAsnGluPheT 1565

    3419 AAACCCGGCCGGCTACC 3435
           |||:::|||::::::
    1565 hrThrLysProSerSer 1570

seq_name: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-460-269C-2

seq_documentation_block:
; Sequence 2, Application US/08460269C
; Patent No. 6197548
;   GENERAL INFORMATION:
;        APPLICANT: CLARE, JEFFREY J.
;                   ROMANOS, MICHAEL A.
;        TITLE OF INVENTION: EXPRESSION OF HETEROLOGOUS PROTEIN IN
;                            YEAST
;        NUMBER OF SEQUENCES: 17
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Millen, White, Zelano & Branigan, P.C.
;             STREET: 2200 Clarendon Blvd., Suite 1400
;             CITY: ARLINGTON
;             STATE: VA
;             COUNTRY: USA
;             ZIP: 22201
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/08/460,269C
;             FILING DATE: 02-Jun-1995
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Lebovitz, Richard M.
;             REGISTRATION NUMBER: 37,067
;             REFERENCE/DOCKET NUMBER: Popov-2
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (703) 243-6333
;             TELEFAX: (703) 243-6410
;   INFORMATION FOR SEQ ID NO: 2:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 910 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-08-460-269C-2


alignment_scores:
            Quality:  283.00              Length:   1100
              Ratio:  0.581                 Gaps:     52
 Percent Similarity:  44.273   Percent Identity: 21.909

alignment_block:
 US-09-303-518D-649 x US-08-460-269C-2   ..

 Align seg 1/1  to: US-08-460-269C-2  from: 1  to: 910

    1366 AACGACCGCCTGTCCAAAATCGGCAAAGGCACGCTGCACGTTCAAGCCAA 1415
         |||:::   :::   |||   |||:::            ::::::   ::
      38 AsnGlnSerIleValLysThrGlyGluArgGlnHisGlyIleHisIleGl 54

    1416 AGGGGAAAACCAAGGCTCGATCAGCGTGGGCGACGGTACAGTCATTTTGG 1465
         :|||::::::   |||::::::   :::::::::||||||:::|||
      54 nGlySerAspProGlyGlyValArgThrAlaSerGlyThrThrIle.... 69

    1466 ATCAGCAGGCAGACGATAAAGGCAAAAAACAAGCCTTTAGTGAAATCGGC 1515

      69 ..................................................... 69

    1516 TTGGTCAGCGGCAGGGGTACGGTGCAACTGAATGCCGATAATCAGTTCAA 1565
         ||||||||||||   :::      :::      :::|||
      70 LysValSerGlyArgGlnAlaGlnGlyIleLeuLeuGluAsn........ 83

    1566 CCCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAACG 1615
                       :::   ||||||:::|||   :::      :::|
      84 ....ProAlaAlaGluLeuGlnPheArgAsnGlySerValThrSerSerG 99

    1616 GGCATTCGCTTTCGTTCCACCGTATTCAAAATACCGATGAAGGGGCGATG 1665
         ||                        |||   :::|||:::|||
      99 ly.......................GlnLeuSerAspAspGlyIle... 106

    1666 ATTGTCAACCACAATCAAGACAAAGAATCCACCGTTACCATTACAGGCAA 1715
                       :::         :::||||||||||:::   :::::
     107 ..............ArgArgPheLeuGlyThrValThrValLysAlaGl 118

    1716 TAAA.........GATATTGCTACAACCGGCAATAACAACAGCTTGGATA 1756
         :|||         |||   ||||||   :::|||   :::
     118 yLysLeuValAlaAspHisAlaThrLeuAlaAsnValGly......... 131

    1757 GCAAAAAAGAAATTGCCTACAACGGTTGGTTTGGCGAGAAAGATACGACC 1806
                                                 ||||||
     132 .........................................AspThrTrp 134

    1807 AAAACGAACGGGCGGCTCAACCTTGTT............TACCAGCCCGC 1844
               :::|||              |||                |||   ::
     135 AspAspAspGlyIleAlaLeuTyrValAlaGlyGluGlnAlaGlnAlaSe 151

    1845 CGCAGAAGACCGCACCCTGCTGCTTTCCGGCGGAACAAATTTA......A 1888
         :      |||   ||||||      :::||||||:::::::::      :
     151 rIleAlaAspSerThrLeuGlnGlyAlaGlyGlyValGlnIleGluArgG 168

    1889 ACGGCAACATCACGCAAACAAACGGCAAACTGTTTTTCAGCGGCAGACCA 1938
         :::::|||:::|||      ::::::   :::      :::|||
     168 lyAlaAsnValThrValGlnArgSerAlaIleValAspGlyGlyLeu... 183

    1939 ACACCGCACGCCTACAATCATTTAAACGACCATTGGTCG...CAAAAAGA 1985
                          |||::::::         |||   |||   ||
     184 ..................HisIleGlyAlaLeuGlnSerLeuGlnProGl 194

    1986 GGGCATTCCTCGCGGGGAAATCGTGTGGGACAACGACTGGATCAACCGCA 2035
         |   :::|||   ::::::::::|||                      |||
     194 uAspLeuProProSerArgValValLeu..................ArgA 205

    2036 CATTTAAAGCGGAAAACTTCCAAATTAAAGGCGGACAGGCGGTGGTTTCC 2085
               ::::::               :::|||:::   |||:::||||||
```



```
 3685 GCCTACCGCCAACAAACCGACCTGCGCCAAATCGGTATGCAGAAAAACCT 3734
       :::   |||::::::         ::::::|||   :::      ::
  667 ArgPheAspGlnLysValAlaGlyPheGluLeuGlyAlaAspHisAlaVa 683

 3735 C............GGCAGCGGGCGCGTCGGCATCCTGTTTTCGCACAACC 3772
      :            |||      ::::::|||   |||   :::::::::|
  683 lAlaValAlaGlyGlyArgTrpHisLeuGlyGlyLeuAlaGlyTyrThrA 700

 3773 GGACCGAAAACACCTTCGACGACGGCATCGGCAACTCGGCACGGCTTGCC 3822
      ||   ::::::   |||          |||:::   :::      :::
  700 rgGlyAspArgGlyPheThrGlyAspGlyGlyGlyHisThrAspSerVal 716

 3823 CACGGCGCCGTTTTCGGGCAATACGGCATCGACAGG...TTCTACATCGG 3869
      |||   :::   ::::::   |||      |||      |||||||:::
  717 HisValGlyGlyTyrAlaThrTyrIleAlaAspSerGlyPheTyrLeuAs 733

 3870 CATCAGCGCGGGCGCGGGTTTTAGCAGCGGCAGCCTT............T 3907
          :::      |||:::      ::::::::::::           |
  733 pAlaThrLeuArgAlaSerArgLeuGluAsnAspPheLysValAlaGlyS 750

 3908 CAGACGGC......ATCGGAGGCAAAATCCGCCGCCGCGTGCTGCATTAC 3951
      ||||||||      :::   ||||||   |||              :::
  750 erAspGlyTyrAlaValLysGlyLysTyrArgThr............His 762

 3952 GGCATTCAGGCACGATACCGCGCCGGT..............TTCGGCGG 3986
      |||:::   |||      :::|||||||                  ||
  763 GlyValGlyAlaSerLeuGluAlaGlyArgArgPheThrHisAlaAspGl 779

 3987 ATTCGGCATCGAACCGCACATCGGCGCAACGCGCTATTTCGTCCAAAAAG 4036
      |:::   :::||||||:::         :::   :::   :::      :
  779 yTrpPheLeuGluProGlnAlaGluLeuAlaValPheArgAlaGlyGlyG 796

 4037 CGGATTACCGCTACGAAAAC...GTCAATATCGCCACCCCCGGCCTTGCA 4083
      ::   ||||||      |||   :::::::::         |||   :::
  796 lyAlaTyrArgAlaAlaAsnGlyLeuArgValArgAspGluGlyGlySer 812

 4084 TTCAACCGCTACCGCGCGGGCATTAAGGCAGATTATTCATTCAAACCGGC 4133
                |||   |||:::::::::         ::::::   ||
  813 SerValLeuGlyArgLeuGlyLeuGluValGlyLysArgIleGluLeuAl 829

 4134 GCAACACATTTCCATCACGCCTTATTTGAGCCTGTCCTATACCGATGCCG 4183
      |         ::::::   ||||||:::::   |||      :::
  829 aGlyGlyArgGlnValGlnProTyrIleLysAlaSerValLeuGlnGluP 846

 4184 CTTCGGGCAAAGTCCGAACACGCGTCAATACCGCCGTATTGGCTCAGGAT 4233
        :::|||         :::::::::|||      :::         :::
  846 heAspGlyAlaGlyThrValHisThrAsnGlyIleAlaHisArgThrGlu 862

 4234 TTCGGCAAAACCCGCAGTGCGGAATGGGGCGTAAACGCCGAAATCAAAGG 4283
      :::      ||||||   ||||||   |||
  863 LeuArgGlyThrArg...AlaGluLeuGly..................... 871

 4284 TTTCACGCTGTCCCTCCACGCTGCCGCCGCCAAAGGCCCGCAACTGGAAG 4333
             |||::::::   ||||||   ::::::|||   :::|||   |
  872 .......LeuGlyMetAlaAlaAlaLeuGlyArgGlyHisSerLeuTyrA 886

 4334 CG......CAACACAGCGCGGGCATCAAATTAGGCTACCGCTGG 4371
      ||      ::::::|||   |||   ||||||:::      |||
  886 laSerTyrGluTyrSerLysGlyProLysLeuAlaMetProTrp 900

seq_name: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-728-470-9

seq_documentation_block:
; Sequence 9, Application US/08728470
; Patent No. 5928651
;  GENERAL INFORMATION:
;    APPLICANT:  Barenkamp, Stephen J
;    TITLE OF INVENTION:  High Molecular Weight Surface Proteins
;    TITLE OF INVENTION:  of No. 5928651-Typeable Haemophilus
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Shoemaker and Mattare, Ltd.
;      STREET:  2001 Jefferson Davis Hwy., 1203 Crystal Plaza
;      STREET:  Bldg. 1
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202-0286
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/728,470
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/302,832
;      FILING DATE:  16-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US PCT/US93/02166
;      FILING DATE:  16-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9205704.1
;      FILING DATE:  16-MAR-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Berkstresser, Jerry W
;      REGISTRATION NUMBER:  22,651
;      REFERENCE/DOCKET NUMBER:  1038-633
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 415-0810
;      TELEFAX:  (703) 415-0813
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1338 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-728-470-9


alignment_scores:
            Quality:  281.00            Length:   1478
              Ratio:  0.404               Gaps:     72
 Percent Similarity:  47.023   Percent Identity: 20.162

alignment_block:
 US-09-303-518D-649 x US-08-728-470-9   ..

 Align seg 1/1  to: US-08-728-470-9  from: 1  to: 1338

  127 GGACACACTTATTTCGGCATCAACTACCAATACTATCGCGACTTTGCCGA 176
      |||:::||||||:::|||                          :::::
   89 GlyGluThrTyrLeuGly.........................GlyAs 96

  177 AAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACA 226
      :::::::|||:::         |||   :::   |||::::::   :::|
   96 pGluArgGlyGlu.........GlyLysAsnGlyIleGlnLeuAlaLysL 110

  227 AAAAAGGGGAGTTGGTCGGCAAATCAATGACAAAAGCCCCGATGATTGAT 276
      ||                    ::::::   |||:::      |||:::
  110 ys..................ThrThrLeuGluLysGlySerThrIleAsn 120

  277 TTTTCTGTGGTGTCGCGTAACGGCGTGGCGGCATTGGTGGGCGATCAATA 326
         |||      ::::::::::|||   |||   :::   ||||||
  121 ValSerGlyLysGluLysGlyGlyArgAlaIleValTrpGlyAspIleAl 137

  327 TATTGTG.........................AGCGTGGCACATAACG 349
       ::::::                        ::::::|||   :::|
  137 aLeuIleAspGlyAsnIleAsnAlaGlnGlyLysAspIleAlaLysThrG 154

  350 GCGGCTAT..................................... 357
      |||||:::
  154 lyGlyPheValGluThrSerGlyHisTyrLeuSerIleAspAspAsnAla 170
```



```
 677 AlaLysPheLysAspIleAsnAsnThrSerSer...LeuAsnIleThrTh 692

1836 CCAGCCCGCCGCAGAAGACCGCACCCTGCTGCTTTCCGGCGGAACAAATT 1885
  :::      :::      ||||||:::                    :
 692 rAsnSerAspThrThrTyrArgThrIle.....................I 702

1886 TAAACGGCAACATCACGCAAACAAACGGCAAACTGTTTTTCAGCGGCAGA 1935
     :::::|||||||||:::::   :::|||   |||   :::      :::
 702 leLysGlyAsnIleSerAsnLysSerGlyAspLeuAsnIleIleAspLys 718

1936 CCAACACCGCACGCCTACAATCATTTAAACGACCATTGGTCGCAAAAAGA 1985
       :::      |||      :::::::::   :::   ||||||||||
 719 LysSer...AspAlaGluIleGlnIleGlyGlyAsnIleSerGlnLysGl 734

1986 GGGCATTCCTCGCGGGGAAATCGTGTGGGACAACGACTGGATCAACCGCA 2035
     |           |||:::::::::   :::::::|||   :::|||   |
 734 u............GlyAsnLeuThrIleSerSerAspLysValAsnIleT 747

2036 CA..............................TTTAAAGCGGAAAAC 2052
     ||                                ::::::::::::
 747 hrAsnGlnIleThrIleLysAlaGlyValGluGlyGlyArgSerAspSer 763

2053 TTCCAAATTAAAGGCGGACAGGCGGTGGTTTCCCGCAATGTTGCCAAAGT 2102
       :::   ::::::::::::   :::::::::   :::      |||::
 764 SerGluAlaGluAsnAlaAsnLeuThrIleGlnThrLysGluLeuLysLe 780

2103 GAAAGGCGATTGGCATTTGAGCAATCACGCCCAAGCAGTTTTTGGTGTCG 2152
     :   ||||||   ::::::|||:::      :::|||   :::   :::
 780 uAlaGlyAspLeuAsnIleSerGlyPheAsnLysAlaGluIleThrAlaL 797

2153 CACCGCATCAAAGCCACACAATCTGTACACGTTCGGACTGGACGGGTCTG 2202
            ::::::   ||||||   :::   |||
 797 ysAsnGlySerAspLeuThrIleGlyAsnAlaSerGly............ 809

2203 ACAAATTGTGTCGAAAAAACCATTACCGACGATAAAGTGATTGCTTCATT 2252
       |||:::      |||   :::|||   |||||||||      |||
 810 GlyAsnAlaAspAlaLysLysValThrPheAspLysValLysAspSerLy 826

2253 GACTAAGACCGACATCAGCGGCAATGTCGATCTTGCCGATCACGCTCATT 2302
       :::||||||   :::   ||||||   |||   ::::::
 826 sIleSerThrAsp...GlyHisAsnValThrLeuAsnSerGlu....... 839

2303 TAAATCTCACAGGGCTTGCCACACTCAACGGCAATCTTAGTGCAAATGGC 2352
                  :::   |||   |||||||:::   :::|||::::::
 840 ..............ValLysThrSerAsnGlySerSerAsnAlaGlyAsn 851

2353 GAT......ACACGTTATACAGTCAGC...CACAACGCCACCCAAAACGG 2393
     |||      |||      |||:::|||      ::::::|||   |||::
 852 AspAsnSerThrGlyLeuThrIleSerAlaLysAspValThrValAsnAs 868

2394 CAACCTTAGCCTCGTGGGCAATGCCCAAGCAACATTTAATCAAGCCACAT 2443
     :|||::::::            ::::::   |||:::|||   ::::::
 868 nAsnValThr............SerHisLysThrIleAsnIleSerAlaA 881

2444 TAAACGGCAACACATCGGCTTCGGGCAATGCTTCATTTAATCTAAGCGAC 2493
        ||||||:::::::::::   :::::::::::::|||   :::
 881 laAlaGlyAsnValThrThrLysGluGlyThrThrIleAsnAlaThrThr 897

2494 CACGCCGTA............CAAAACGGCAGTCTGACGCTTTCCGGCAA 2531
       :::|||            |||||||||      |||::::::|||||
 898 GlySerValGluValThrAlaGlnAsnGly......ThrIleLysGlyAs 912

2532 C...GCTAAGGCAAACGTAAGCCATTCCGCACTCAACGGTAATGTCTCCC 2578
     |   ::::::   ||||||:::   :::|||   :::   |||:::
 912 nIleThrSerGlnAsnValThrValThrAlaThrGlu...AsnLeuValT 928

2579 TAGCCGATAAGGCAGTATTCCATTTTGAAAGC..............AGC 2613
       :::::::::|||||||::::::      |||              |||
 928 hrThrGluAsnAlaValIleAsnAlaThrSerGlyThrValAsnIleSer 944

2614 CGCTTTACCGGACAAATCAGCGGCGGCAAGGATACGGCA.........TT 2654
          ||||||:::|||:::||||||   :::::::::         ::
 945 ThrLysThrGlyAspIleLysGlyGlyIleGluSerThrSerGlyAsnVa 961
```

```
2655 ACACTTAAAAGACAGCGAATGGACGCTGCCGTCAGGCACGGAATTAGGCA 2704
     :::::::      |||      ||||||            :::::::::|
  961 lAsnIleThrAlaSerGlyAsnThrLeu...........LysValSerA 974

2705 ATTTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCGCCAC 2754
     ||::::::   :::::::::|||:::|||   :::
  974 snIleThrGlyGlnAspValThrValThrAlaAsp.............. 985

2755 GATGCGGCAGGGGCGCAAACCGGCAGTGCGACAGATGCGCCGCGCCGCCG 2804
            |||||||||   |||   :::|||
  986 ......AlaGlyAlaLeuThrThrThrAla..................... 993

2805 TTCGCGCCGTTCGCGCCGTTCCCTATTATCCGTTACACCGCCAACTTCGG 2854
                        |||   :::|||:::|||      ::::::
  994 ................GlySerThrIleSerAlaThrThrGlyAsnAla. 1004

2855 TAGAATCCCGTTTCAACACGCTGACGGTAAACGGCAAATTGAACGGTCAG 2904
                  |||      |||   :::|||   :::|||||||:::
 1005 .............AsnIleThrThrLysThrGlyAspIleAsnGlyLys 1016

2905 GGAACATTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGACAAATT 2954

1016 .................................................. 1016

2955 GAAGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCGGTCAACAAT. 3003
            :::|||||||||:::|||:::   ||||||:::::::::::::
1017 .......ValGluSerSerSerGlySerValThrLeuValAlaThrGlyA 1031

3004 ...........ACCGGCAACGAACCTGCAAGCCTCGAACAATTGACGGTA 3042
                  :::|||||||      ::::::          :::|||:::
1031 laThrLeuAlaValGlyAsnIleSerGlyAsnThrValThrIleThrAla 1047

3043 GTGGAAGGAAAAGACAACAAACCGCTGTCCGAAAACCTTAAT........ 3084
        :::||||||   ::::::   :::::::::::::::|||
1048 AspSerGlyLysLeuThrSerThrValGlySerThrIleAsnGlyThrAs 1064

3085 ..................................TTCACCCTGCAAA 3097
                                         |||::::::::
1064 nSerValThrThrSerSerGlnSerGlyAspIleGluGlyThrIleSerG 1081

3098 ACGAACACGTCGATGCCGGCGCGTGGCGTTACCAACTCATCCGCAAAGAC 3147
     :::::   |||::::::   |||                    :::
1081 lyAsnThrValAsnValThrAla.....................SerThr 1090

3148 GGCGAGTTCCGCCTGCATAATCCGGTCAAAGAACAAGAGCTTTCCGACAA 3197
     |||::::::   :::   |||   :::|||          ::::::   ||
1091 GlyAspLeuThrIleGlyAsnSerAlaLys.........ValGluAlaLy 1104

3198 ACTCGGCAAGGCAGAAGCCAAAAAACAGGCGGAAAAAGACAACGCGCAAA 3247
     |   |||   |||              ::::::   |||   ::::::|||:
1104 sAsnGlyAlaAlaThrLeuThrAlaGluSerGlyLysLeuThrThrGlnT 1121

3248 GCCTTGACGCGCTGATTGCGGCCGGGCGCGATGCCGTCGAAAAGACAGAA 3297
     ::   :::::::::   :::::::::               :::|||
1121 hrGlySerSerIleThrSerSerAsn............GlyGlnThrThr 1133

3298 AGCGTTGCCGAACCGGCCCGGCAGGCAGGCGGGGAAAAT........... 3336
        :::|||:::   :::      ||||||:::   |||
1134 LeuThrAlaLysAspSerSerIleAlaGlyAsnIleAsnAlaAlaAsnVa 1150

3337 .............GTCGGCATTATGCAGGCGGAGGAAGAGAAAAAACGGG 3373
                  :::|||   :::   :::      ::::::|||   :
1150 lThrLeuAsnThrThrGlyThrLeuThrThrThrGlyAspSerLys...I 1166

3374 TGCAGGCGGATAAAGACACCGCCTTGGCGAAACAGCGCGAAGCGGAAACC 3423
     :::::|||   :::   |||        :::   ::::::|||:::
1166 leAsnAlaThrSerGlyThrLeuThrIleAsnAlaLysAspAlaLysLeu 1182

3424 CGGCCGGCTACCACCGCCTTCCCCCGCGCCCGCCGCGCCCGCCGGGATTT 3473
          |||:::::::::
1183 AspGlyAlaAlaSerGly............................... 1188
```

```
    3474 GCCGCAACTGCAACCCCAACCGCAGCCCCAACCGCAGCGCGACCTGATCA 3523
                                            :::|||   :::::::
    1189 ..............................AspArgThrValValA 1194

    3524 GCCGTTATGCCAATAGCGGTTTGAGTGAATTTTCCGCCACGCTCAACAGC 3573
         ::          ||||||   ::::::   :::|||      :::|||
    1194 snAlaThrAsnAlaSerGlySerGlyAsnValThrAlaLysThrSerSer 1210

    3574 GTTTTCGCCGTACAGGACGAATTAGAC..............CGCGTATT 3608
                 :::      :::|||:::                ::::::::
    1211 SerValAsnIleThrGlyAspLeuAsnThrIleAsnGlyLeuAsnIleIl 1227

    3609 TGCCGAAGACCGCCGCAACGCCGTTTGGACAAGCGGCATCCGG...GACA 3655
         ::::|||:::   ||||||:::|||         |||   :::   |||:
    1227 eSerGluAsnGlyArgAsnThrValArgLeuArgGlyLysGluIleAspV 1244

    3656 CCAAACACTACCGTTCGCAAGATTTCCGCGCCTACCGCCAACAAACCGAC 3705
         ::|||:::         |||    ::::::   ::::::::::::
    1244 alLysTyrIle......GlnPro.GlyValAlaSerValGluGluValIl 1258

    3706 CTGCGCCAAATCGGTATGCAGAAAAACC..................... 3733
             ||||||   |||   :::|||:::
    1258 eGluAlaLysArgValLeuGluLysValLysAspLeuSerAspGluGluA 1275

    3734 .....TCGGCAGCGGGCGCGTCGGCATCCTGTTTTCGCACAACCGGACCG 3778
              :::   |||      :::::::|||:::   |||   :::   |||
    1275 rgGluThrLeuAlaLysLeuGlyValSerAlaValArgPheValGluPro 1291

    3779 AAAACACCTTCGACGACGGCATCGGCAACTCGGCACGGCTTGCCCACGGC 3828
         ::::::      |||:::         |||::::::   :::   |||
    1292 AsnAsnAlaIleThrVal......AsnThrGlnAsnGluPheThrThrLy 1306

    3829 GCCGTTTTCGGGCAATACGGCATCGACAGGTTCTACA......TCGGCAT 3872
          |||   |||      |||   |||   |||::::::      |||::::
    1306 sProSerSerGlnValThrIleSerGluGlyLysAlaCysPheSerSerG 1323

    3873 CAGCGCGGGCGCGGGTTTTAGCAGCGGCAGCC 3904
         ::      |||||||||   :::   :::|||
    1323 lyAsnGlyAlaArgValCysThrAsnValAla 1333

seq_name: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-719-641-9

seq_documentation_block:
; Sequence 9, Application US/08719641
; Patent No. 6218141
;  GENERAL INFORMATION:
;    APPLICANT:  Barenkamp, Stephen J
;    TITLE OF INVENTION:  High Molecular Weight Surface Proteins
;    TITLE OF INVENTION:  of No. 6218141-Typeable Haemophilus
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Shoemaker and Mattare, Ltd.
;      STREET:  2001 Jefferson Davis Hwy., 1203 Crystal Plaza
;      STREET:  Bldg. 1
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202-0286
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/719,641
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/302,832
;      FILING DATE:  16-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US PCT/US93/02166
;      FILING DATE:  16-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9205704.1
;      FILING DATE:  16-MAR-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Berkstresser, Jerry W
;      REGISTRATION NUMBER:  22,651
;      REFERENCE/DOCKET NUMBER:  1038-625
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 415-0810
;      TELEFAX:  (703) 415-0813
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1338 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-719-641-9


alignment_scores:
           Quality:  281.00              Length:   1478
             Ratio:  0.404                 Gaps:     72
 Percent Similarity:  47.023   Percent Identity: 20.162

alignment_block:
 US-09-303-518D-649 x US-08-719-641-9    ..

 Align seq 1/1  to: US-08-719-641-9  from: 1  to: 1338

     127 GGACACACTTATTTCGGCATCAACTACCAATACTATCGCGACTTTGCCGA 176
         |||:::||||||:::|||                              :::::
      89 GlyGluThrTyrLeuGly..........................GlyAs 96

     177 AAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACA 226
         :::::::|||:::         |||   :::   |||:::::   :::|
      96 pGluArgGlyGlu.........GlyLysAsnGlyIleGlnLeuAlaLysL 110

     227 AAAAAGGGGAGTTGGTCGGCAAATCAATGACAAAAGCCCCGATGATTGAT 276
         ||                  ::::::   |||:::      |||:::
     110 ys.................ThrThrLeuGluLysGlySerThrIleAsn 120

     277 TTTTCTGTGGTGTCGCGTAACGGCGTGGCGGCATTGGTGGGCGATCAATA 326
            |||      ::::::::::|||   |||   :::   ||||||
     121 ValSerGlyLysGluLysGlyGlyArgAlaIleValTrpGlyAspIleAl 137

     327 TATTGTG..........................AGCGTGGCACATAACG 349
          ::::::                         :::::::|||   :::|
     137 aLeuIleAspGlyAsnIleAsnAlaGlnGlyLysAspIleAlaLysThrG 154

     350 GCGGCTAT.......................................... 357
         |||||:::
     154 lyGlyPheValGluThrSerGlyHisTyrLeuSerIleAspAspAsnAla 170

     357 ..... 357

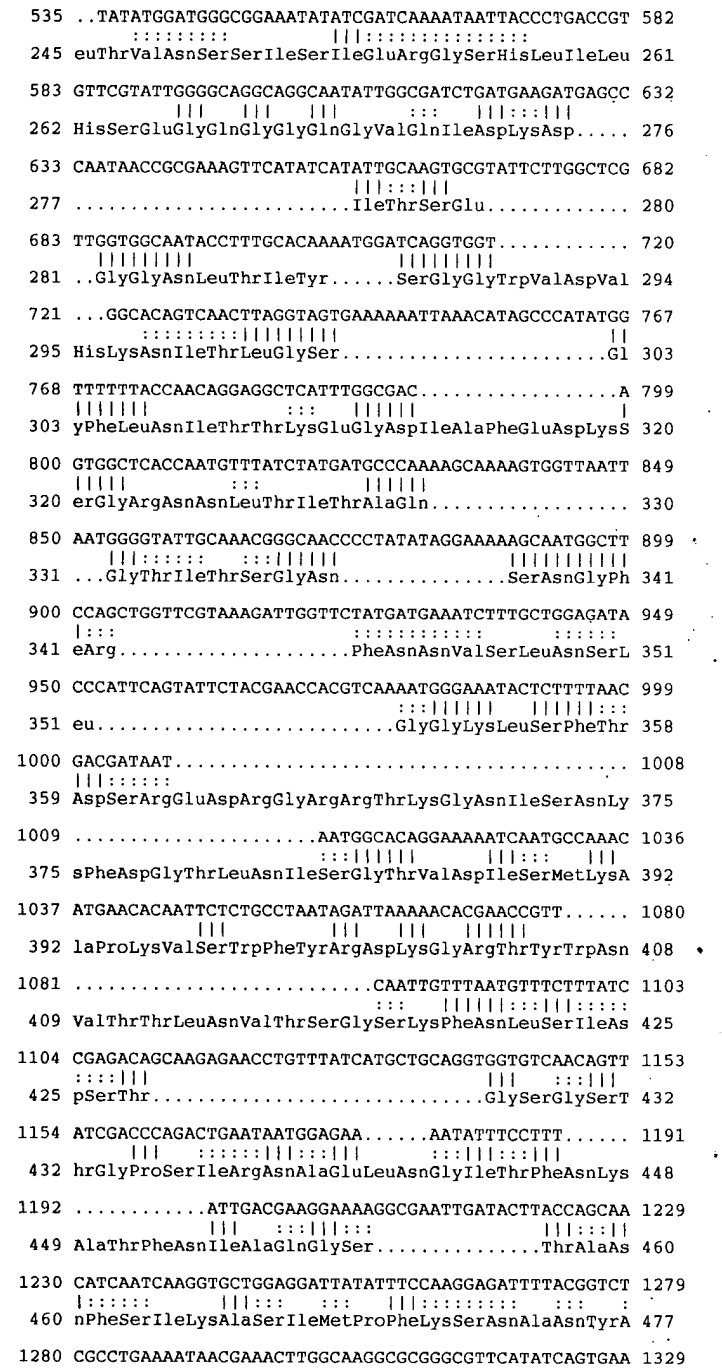
     171 IleValLysThrLysGluTrpLeuLeuAspProGluAsnValThrIleGl 187

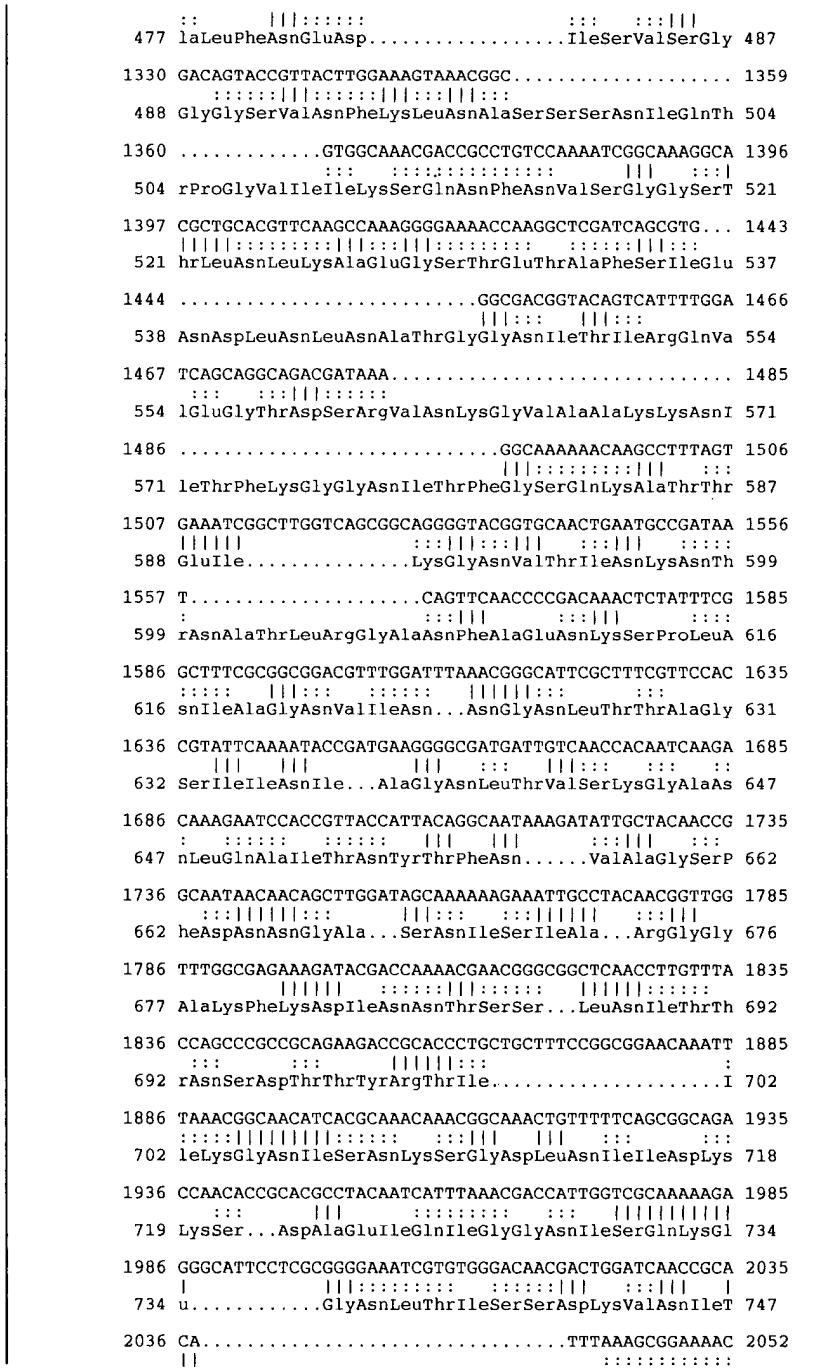
     358 .............AACAACGTTGATTTTGGTGCGGAAGGAAGAAATCCCG 394
                     ::::::|||::::::||||||:::   ||||||
     187 uAlaProSerAlaSerArgValGluLeuGlyAlaAsp...ArgAsnSerH 203

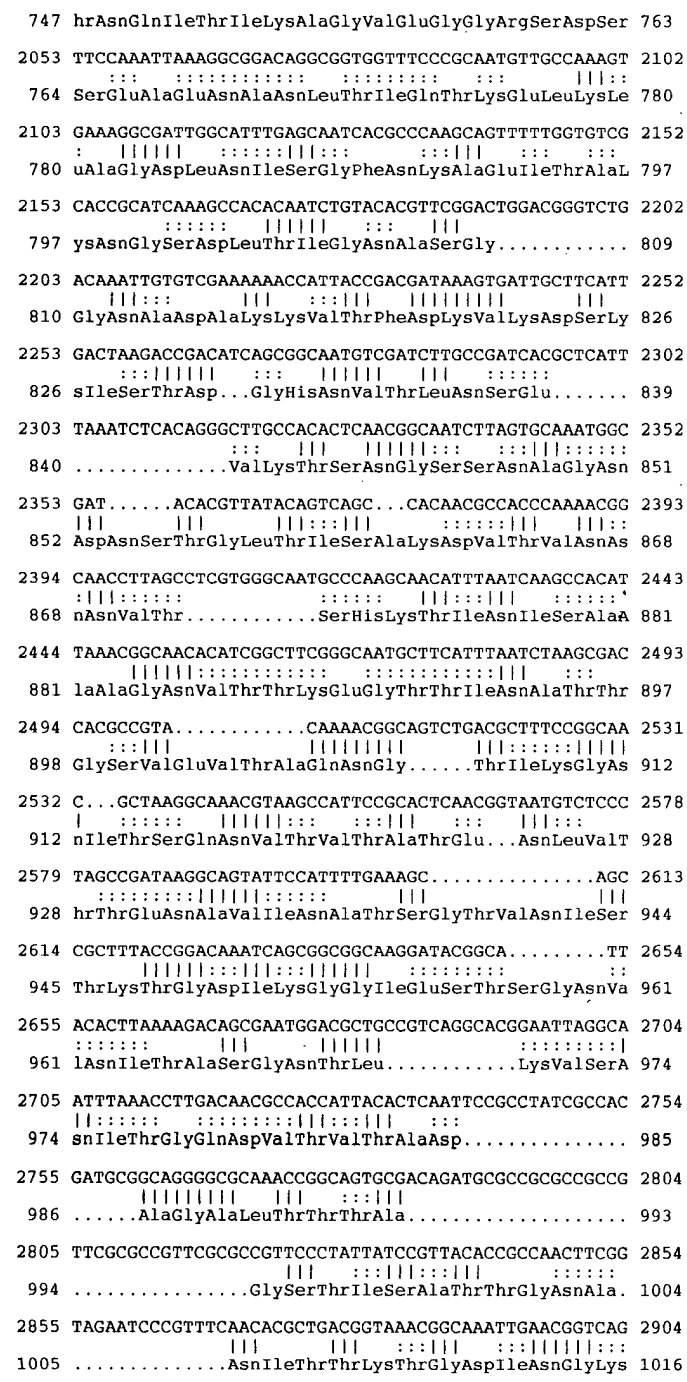
     395 ATCAACATCGTTTTACTTATAAAATTGTGAAACGGAATAATTATAAAGCA 444
           :::   :::              :::|||:::||||||   :::
     203 isSerAlaGluValIleLysValThrLeuLysLysAsnAsnThrSerLeu 219

     445 GGGACTAAAGGCCATCCTTATGGCGGCGATTATCATATGCCGCGTTTGCA 494
            |||      :::                              :::
     220 ThrThrLeuThrAsn.......................ThrThrIleSe 228

     495 TAAATTTGTCACAGATGCAGAACCTGTTGAAATGACCAGT.......... 534
          :::::::::   :::|||:::   |||:::::::|||:::
     228 rAsnLeuLeuLysSerAlaHisValValAsnIleThrAlaArgArgLysL 245
```


535 ..TATATGGATGGGCGGAAATATATCGATCAAAATAATTACCCTGACCGT 582
245 euThrValAsnSerSerIleSerIleGluArgGlySerHisLeuIleLeu 261
583 GTTCGTATTGGGGCAGGCAGGCAATATTGGCGATCTGATGAAGATGAGCC 632
262 HisSerGluGlyGlnGlyGlyGlnGlyValGlnIleAspLysAsp..... 276
633 CAATAACCGCGAAAGTTCATATCATATTGCAAGTGCGTATTCTTGGCTCG 682
277 IleThrSerGlu..... 280
683 TTGGTGGCAATACCTTTGCACAAAATGGATCAGGTGGT..... 720
281 ..GlyGlyAsnLeuThrIleTyr.....SerGlyGlyTrpValAspVal 294
721 ...GGCACAGTCAACTTAGGTAGTGAAAAAATTAAACATAGCCCATATGG 767
295 HisLysAsnIleThrLeuGlySer.....Gl 303
768 TTTTTTACCAACAGGAGGCTCATTTGGCGAC.....A 799
303 yPheLeuAsnIleThrThrLysGluGlyAspIleAlaPheGluAspLysS 320
800 GTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGGTTAATT 849
320 erGlyArgAsnAsnLeuThrIleThrAlaGln..... 330
850 AATGGGGTATTGCAAACGGGCAACCCCTATATAGGAAAAAGCAATGGCTT 899
331 ...GlyThrIleThrSerGlyAsn.....SerAsnGlyPh 341
900 CCAGCTGGTTCGTAAAGATTGGTTCTATGATGAAATCTTTGCTGGAGATA 949
341 eArg.....PheAsnAsnValSerLeuAsnSerL 351
950 CCCATTCAGTATTCTACGAACCACGTCAAAATGGGAAATACTCTTTTAAC 999
351 eu.....GlyGlyLysLeuSerPheThr 358
1000 GACGATAAT..... 1008
359 AspSerArgGluAspArgGlyArgArgThrLysGlyAsnIleSerAsnLy 375
1009 AATGGCACAGGAAAAATCAATGCCAAAC 1036
375 sPheAspGlyThrLeuAsnIleSerGlyThrValAspIleSerMetLysA 392
1037 ATGAACACAATTCTCTGCCTAATAGATTAAAAACACGAACCGTT..... 1080
392 laProLysValSerTrpPheTyrArgAspLysGlyArgThrTyrTrpAsn 408
1081 CAATTGTTTAATGTTTCTTTATC 1103
409 ValThrThrLeuAsnValThrSerGlySerLysPheAsnLeuSerIleAs 425
1104 CGAGACAGCAAGAGAACCTGTTTATCATGCTGCAGGTGGTGTCAACAGTT 1153
425 pSerThr.....GlySerGlySerT 432
1154 ATCGACCCAGACTGAATAATGGAGAA.....AATATTTCCTTT..... 1191
432 hrGlyProSerIleArgAsnAlaGluLeuAsnGlyIleThrPheAsnLys 448
1192 ATTGACGAAGGAAAAGGCGAATTGATACTTACCAGCAA 1229
449 AlaThrPheAsnIleAlaGlnGlySer.....ThrAlaAs 460
1230 CATCAATCAAGGTGCTGGAGGATTATATTTCCAAGGAGATTTTACGGTCT 1279
460 nPheSerIleLysAlaSerIleMetProPheLysSerAsnAlaAsnTyrA 477
1280 CGCCTGAAAATAACGAAACTTGGCAAGGCGCGGGCGTTCATATCAGTGAA 1329


477 laLeuPheAsnGluAsp.....IleSerValSerGly 487
1330 GACAGTACCGTTACTTGGAAAGTAAACGGC..... 1359
488 GlyGlySerValAsnPheLysLeuAsnAlaSerSerSerAsnIleGlnTh 504
1360 GTGGCAAACGACCGCCTGTCCAAAATCGGCAAAGGCA 1396
504 rProGlyValIleIleLysSerGlnAsnPheAsnValSerGlyGlySerT 521
1397 CGCTGCACGTTCAAGCCAAAGGGGAAAACCAAGGCTCGATCAGCGTG... 1443
521 hrLeuAsnLeuLysAlaGluGlySerThrGluThrAlaPheSerIleGlu 537
1444 GGCGACGGTACAGTCATTTTGGA 1466
538 AsnAspLeuAsnLeuAsnAlaThrGlyGlyAsnIleThrIleArgGlnVa 554
1467 TCAGCAGGCAGACGATAAA..... 1485
554 lGluGlyThrAspSerArgValAsnLysGlyValAlaAlaLysLysAsnI 571
1486 GGCAAAAAACAAGCCTTTAGT 1506
571 leThrPheLysGlyGlyAsnIleThrPheGlySerGlnLysAlaThrThr 587
1507 GAAATCGGCTTGGTCAGCGGCAGGGGTACGGTGCAACTGAATGCCGATAA 1556
588 GluIle.....LysGlyAsnValThrIleAsnLysAsnTh 599
1557 T.....CAGTTCAACCCCGACAAACTCTATTTCG 1585
599 rAsnAlaThrLeuArgGlyAlaAsnPheAlaGluAsnLysSerProLeuA 616
1586 GCTTTCGCGGCGGACGTTTGGATTTAAACGGGCATTCGCTTTCGTTCCAC 1635
616 snIleAlaGlyAsnValIleAsn...AsnGlyAsnLeuThrThrAlaGly 631
1636 CGTATTCAAAATACCGATGAAGGGGCGATGATTGTCAACCACAATCAAGA 1685
632 SerIleIleAsnIle...AlaGlyAsnLeuThrValSerLysGlyAlaAs 647
1686 CAAAGAATCCACCGTTACCATTACAGGCAATAAAGATATTGCTACAACCG 1735
647 nLeuGlnAlaIleThrAsnTyrThrPheAsn.....ValAlaGlySerP 662
1736 GCAATAACAACAGCTTGGATAGCAAAAAAGAAATTGCCTACAACGGTTGG 1785
662 heAspAsnAsnGlyAla...SerAsnIleSerIleAla...ArgGlyGly 676
1786 TTTGGCGAGAAAGATACGACCAAAACGAACGGGCGGCTCAACCTTGTTTA 1835
677 AlaLysPheLysAspIleAsnAsnThrSerSer...LeuAsnIleThrTh 692
1836 CCAGCCCGCCGCAGAAGACCGCACCCTGCTGCTTTCCGGCGGAACAAATT 1885
692 rAsnSerAspThrThrTyrArgThrIle.....I 702
1886 TAAACGGCAACATCACGCAAACAAACGGCAAACTGTTTTTCAGCGGCAGA 1935
702 leLysGlyAsnIleSerAsnLysSerGlyAspLeuAsnIleIleAspLys 718
1936 CCAACACCGCACGCCTACAATCATTTAAACGACCATTGGTCGCAAAAAGA 1985
719 LysSer...AspAlaGluIleGlnIleGlyGlyAsnIleSerGlnLysGl 734
1986 GGGCATTCCTCGCGGGGAAATCGTGTGGGACAACGACTGGATCAACCGCA 2035
734 u.....GlyAsnLeuThrIleSerSerAspLysValAsnIleT 747
2036 CA.....TTTAAAGCGGAAAAC 2052


747 hrAsnGlnIleThrIleLysAlaGlyValGluGlyGlyArgSerAspSer 763
2053 TTCCAAATTAAAGGCGGACAGGCGGTGGTTTCCCGCAATGTTGCCAAAGT 2102
764 SerGluAlaGluAsnAlaAsnLeuThrIleGlnThrLysGluLeuLysLe 780
2103 GAAAGGCGATTGGCATTTGAGCAATCACGCCCAAGCAGTTTTTGGTGTCG 2152
780 uAlaGlyAspLeuAsnIleSerGlyPheAsnLysAlaGluIleThrAlaL 797
2153 CACCGCATCAAAGCCACACAATCTGTACACGTTCGGACTGGACGGGTCTG 2202
797 ysAsnGlySerAspLeuThrIleGlyAsnAlaSerGly..... 809
2203 ACAAATTGTGTCGAAAAAACCATTACCGACGATAAAGTGATTGCTTCATT 2252
810 GlyAsnAlaAspAlaLysLysValThrPheAspLysValLysAspSerLy 826
2253 GACTAAGACCGACATCAGCGGCAATGTCGATCTTGCCGATCACGCTCATT 2302
826 sIleSerThrAsp...GlyHisAsnValThrLeuAsnSerGlu..... 839
2303 TAAATCTCACAGGGCTTGCCACACTCAACGGCAATCTTAGTGCAAATGGC 2352
840 ValLysThrSerAsnGlySerSerAsnAlaGlyAsn 851
2353 GAT.....ACACGTTATACAGTCAGC...CACAACGCCACCCAAAACGG 2393
852 AspAsnSerThrGlyLeuThrIleSerAlaLysAspValThrValAsnAs 868
2394 CAACCTTAGCCTCGTGGGCAATGCCCAAGCAACATTTAATCAAGCCACAT 2443
868 nAsnValThr.....SerHisLysThrIleAsnIleSerAlaA 881
2444 TAAACGGCAACACATCGGCTTCGGGCAATGCTTCATTTAATCTAAGCGAC 2493
881 laAlaGlyAsnValThrThrLysGluGlyThrThrIleAsnAlaThrThr 897
2494 CACGCCGTA.....CAAAACGGCAGTCTGACGCTTTCCGGCAA 2531
898 GlySerValGluValThrAlaGlnAsnGly.....ThrIleLysGlyAs 912
2532 C...GCTAAGGCAAACGTAAGCCATTCCGCACTCAACGGTAATGTCTCCC 2578
912 nIleThrSerGlnAsnValThrValThrAlaThrGlu...AsnLeuValT 928
2579 TAGCCGATAAGGCAGTATTCCATTTTGAAAGC.....AGC 2613
928 hrThrGluAsnAlaValIleAsnAlaThrSerGlyThrValAsnIleSer 944
2614 CGCTTTACCGGACAAATCAGCGGCGGCAAGGATACGGCA.....TT 2654
945 ThrLysThrGlyAspIleLysGlyGlyIleGluSerThrSerGlyAsnVa 961
2655 ACACTTAAAAGACAGCGAATGGACGCTGCCGTCAGGCACGGAATTAGGCA 2704
961 lAsnIleThrAlaSerGlyAsnThrLeu.....LysValSerA 974
2705 ATTTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCGCCAC 2754
974 snIleThrGlyGlnAspValThrValThrAlaAsp..... 985
2755 GATGCGGCAGGGGCGCAAACCGGCAGTGCGACAGATGCGCCGCGCCGCCG 2804
986 AlaGlyAlaLeuThrThrThrAla..... 993
2805 TTCGCGCCGTTCGCGCCGTTCCCTATTATCCGTTACACCGCCAACTTCGG 2854
994 GlySerThrIleSerAlaThrThrGlyAsnAla. 1004
2855 TAGAATCCCGTTTCAACACGCTGACGGTAAACGGCAAATTGAACGGTCAG 2904
1005 AsnIleThrThrLysThrGlyAspIleAsnGlyLys 1016
2905 GGAACATTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGACAAATT 2954
1016 1016
2955 GAAGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCGGTCAACAAT. 3003
1017 ValGluSerSerSerGlySerValThrLeuValAlaThrGlyA 1031
3004 ACCGGCAACGAACCTGCAAGCCTCGAACAATTGACGGTA 3042
1031 laThrLeuAlaValGlyAsnIleSerGlyAsnThrValThrIleThrAla 1047
3043 GTGGAAGGAAAAGACAACAAACCGCTGTCCGAAAACCTTAAT..... 3084
1048 AspSerGlyLysLeuThrSerThrValGlySerThrIleAsnGlyThrAs 1064
3085 TTCACCCTGCAAA 3097
1064 nSerValThrThrSerSerGlnSerGlyAspIleGluGlyThrIleSerG 1081
3098 ACGAACACGTCGATGCCGGCGCGTGGCGTTACCAACTCATCCGCAAAGAC 3147
1081 lyAsnThrValAsnValThrAla.....SerThr 1090
3148 GGCGAGTTCCGCCTGCATAATCCGGTCAAAGAACAAGAGCTTTCCGACAA 3197
1091 GlyAspLeuThrIleGlyAsnSerAlaLys.....ValGluAlaLy 1104
3198 ACTCGGCAAGGCAGAAGCCAAAAAACAGGCGGAAAAAGACAACGCGCAAA 3247
1104 sAsnGlyAlaAlaThrLeuThrAlaGluSerGlyLysLeuThrThrGlnT 1121
3248 GCCTTGACGCGCTGATTGCGGCCGGGCGCGATGCCGTCGAAAAGACAGAA 3297
1121 hrGlySerSerIleThrSerSerAsn.....GlyGlnThrThr 1133
3298 AGCGTTGCCGAACCGGCCCGGCAGGCAGGCGGGGAAAAT..... 3336
1134 LeuThrAlaLysAspSerSerIleAlaGlyAsnIleAsnAlaAlaAsnVa 1150
3337 GTCGGCATTATGCAGGCGGAGGAAGAGAAAAAACGGG 3373
1150 lThrLeuAsnThrThrGlyThrLeuThrThrThrGlyAspSerLys...I 1166
3374 TGCAGGCGGATAAAGACACCGCCTTGGCGAAACAGCGCGAAGCGGAAACC 3423
1166 leAsnAlaThrSerGlyThrLeuThrIleAsnAlaLysAspAlaLysLeu 1182
3424 CGGCCGGCTACCACCGCCTTCCCCCGCGCCCGCCGCGCCCGCCGGGATTT 3473
1183 AspGlyAlaAlaSerGly..... 1188
3474 GCCGCAACTGCAACCCCAACCGCAGCCCCAACCGCAGCGCGACCTGATCA 3523
1189 AspArgThrValValA 1194
3524 GCCGTTATGCCAATAGCGGTTTGAGTGAATTTTCCGCCACGCTCAACAGC 3573
1194 snAlaThrAsnAlaSerGlySerGlyAsnValThrAlaLysThrSerSer 1210
3574 GTTTTCGCCGTACAGGACGAATTAGAC.....CGCGTATT 3608
1211 SerValAsnIleThrGlyAspLeuAsnThrIleAsnGlyLeuAsnIleIl 1227
3609 TGCCGAAGACCGCCGCAACGCCGTTTGGACAAGCGGCATCCGG...GACA 3655
1227 eSerGluAsnGlyArgAsnThrValArgLeuArgGlyLysGluIleAspV 1244
3656 CCAAACACTACCGTTCGCAAGATTTCCGCGCCTACCGCCAACAAACCGAC 3705
1244 alLysTyrIle.....GlnPro.GlyValAlaSerValGluGluValIl 1258

```
     3706 CTGCGCCAAATCGGTATGCAGAAAAACC..................... 3733
          ||||||   |||   :::|||:::
     1258 eGluAlaLysArgValLeuGluLysValLysAspLeuSerAspGluGluA 1275

     3734 .....TCGGCAGCGGGCGCGTCGGCATCCTGTTTTCGCACAACCGGACCG 3778
               :::   |||      ::::::|||:::   |||   :::   |||
     1275 rgGluThrLeuAlaLysLeuGlyValSerAlaValArgPheValGluPro 1291

     3779 AAAACACCTTCGACGACGGCATCGGCAACTCGGCACGGCTTGCCCACGGC 3828
          ::::::      |||:::         |||::::::   :::   |||
     1292 AsnAsnAlaIleThrVal......AsnThrGlnAsnGluPheThrThrLy 1306

     3829 GCCGTTTTCGGGCAATACGGCATCGACAGGTTCTACA......TCGGCAT 3872
           |||   |||      |||   |||   |||::::::      |||::::
     1306 sProSerSerGlnValThrIleSerGluGlyLysAlaCysPheSerSerG 1323

     3873 CAGCGCGGGCGCGGGTTTTAGCAGCGGCAGCC 3904
          ::      |||||||||   :::   :::|||
     1323 lyAsnGlyAlaArgValCysThrAsnValAla 1333

seq_name: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-617-697-9

seq_documentation_block:
; Sequence 9, Application US/08617697
; Patent No. 5977336
;  GENERAL INFORMATION:
;    APPLICANT:  Barenkamp, Stephen J
;    TITLE OF INVENTION:  High Molecular Weight Surface Proteins
;    TITLE OF INVENTION:  of No. 5977336-Typeable Haemophilus
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Shoemaker and Mattare, Ltd.
;      STREET:  2001 Jefferson Davis Hwy., 1203 Crystal Plaza
;      STREET:  Bldg. 1
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202-0286
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/617,697
;      FILING DATE:  01-APR-1996
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/302,832
;      FILING DATE:  05-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US PCT/US93/02166
;      FILING DATE:  16-MAR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Berkstresser, Jerry W
;      REGISTRATION NUMBER:  22,651
;      REFERENCE/DOCKET NUMBER:  1038-557
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 415-0810
;      TELEFAX:  (703) 415-0813
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1599 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-617-697-9


alignment_scores:
           Quality:  281.00              Length:   1478
             Ratio:  0.404                Gaps:     72
 Percent Similarity:  47.023   Percent Identity: 20.162
```

```
alignment_block:
 US-09-303-518D-649 x US-08-617-697-9   ..

 Align seg 1/1  to: US-08-617-697-9  from: 1  to: 1599

     127 GGACACACTTATTTCGGCATCAACTACCAATACTATCGCGACTTTGCCGA 176
         |||:::||||||:::|||                              :::::
     349 GlyGluThrTyrLeuGly.........................GlyAs 356

     177 AAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACA 226
         :::::::|||:::          |||   :::   |||::::::   :::|
     356 pGluArgGlyGlu.........GlyLysAsnGlyIleGlnLeuAlaLysL 370

     227 AAAAAGGGGAGTTGGTCGGCAAATCAATGACAAAAGCCCCGATGATTGAT 276
         ||                    ::::::   |||:::      |||:::
     370 ys..................ThrThrLeuGluLysGlySerThrIleAsn 380

     277 TTTTCTGTGGTGTCGCGTAACGGCGTGGCGGCATTGGTGGGCGATCAATA 326
            |||      :::::::::|||   |||   :::   ||||||
     381 ValSerGlyLysGluLysGlyGlyArgAlaIleValTrpGlyAspIleAl 397

     327 TATTGTG.........................AGCGTGGCACATAACG 349
          ::::::                        ::::::|||   :::|
     397 aLeuIleAspGlyAsnIleAsnAlaGlnGlyLysAspIleAlaLysThrG 414

     350 GCGGCTAT..................................... 357
         |||||:::
     414 lyGlyPheValGluThrSerGlyHisTyrLeuSerIleAspAspAsnAla 430

     357 ..... 357

     431 IleValLysThrLysGluTrpLeuLeuAspProGluAsnValThrIleGl 447

     358 .............AACAACGTTGATTTTGGTGCGGAAGGAAGAAATCCCG 394
                     ::::::|||::::::||||||:::   ||||||
     447 uAlaProSerAlaSerArgValGluLeuGlyAlaAsp...ArgAsnSerH 463

     395 ATCAACATCGTTTTACTTATAAAATTGTGAAACGGAATAATTATAAAGCA 444
            :::   :::              :::|||:::||||||   :::
     463 isSerAlaGluValIleLysValThrLeuLysLysAsnAsnThrSerLeu 479

     445 GGGACTAAAGGCCATCCTTATGGCGGCGATTATCATATGCCGCGTTTGCA 494
            |||      :::                              :::
     480 ThrThrLeuThrAsn.......................ThrThrIleSe 488

     495 TAAATTTGTCACAGATGCAGAACCTGTTGAAATGACCAGT.......... 534
          :::::::::   :::|||:::   |||::::::|||:::
     488 rAsnLeuLeuLysSerAlaHisValValAsnIleThrAlaArgArgLysL 505

     535 ..TATATGGATGGGCGGAAATATATCGATCAAAATAATTACCCTGACCGT 582
               :::::::::          |||:::::::::::::::
     505 euThrValAsnSerSerIleSerIleGluArgGlySerHisLeuIleLeu 521

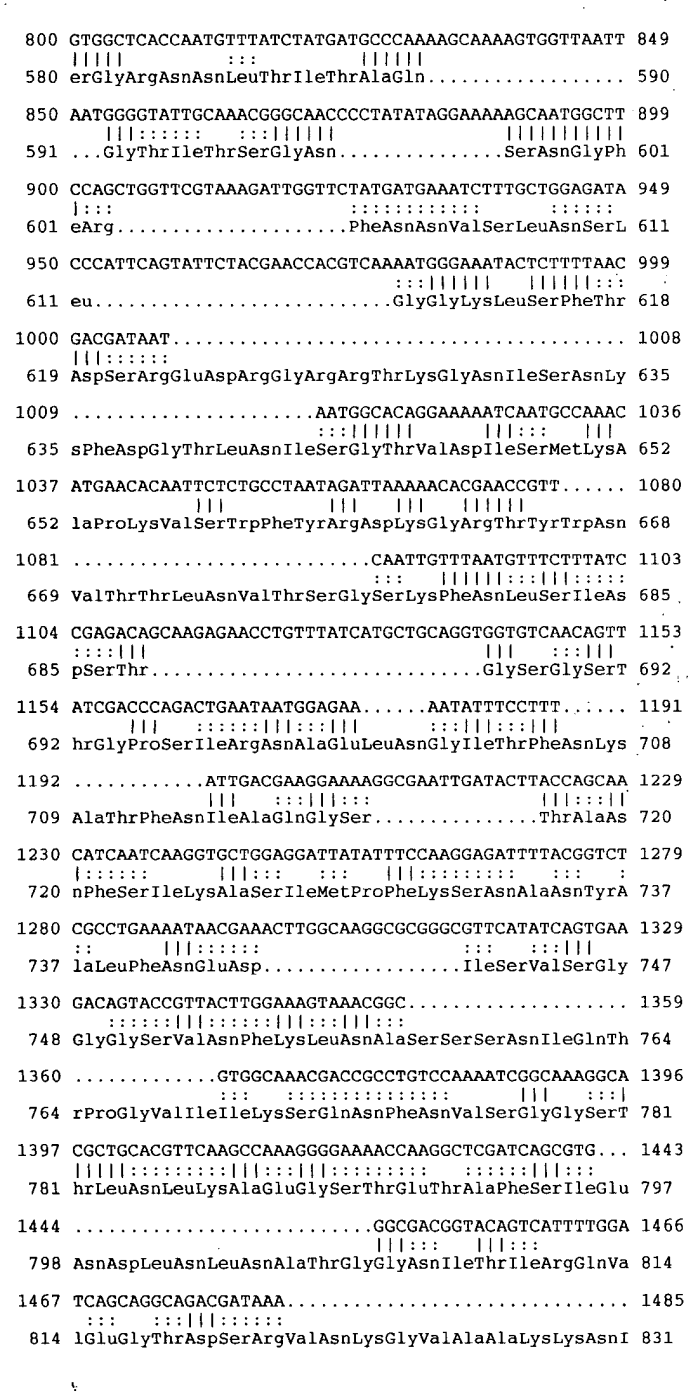
     583 GTTCGTATTGGGGCAGGCAGGCAATATTGGCGATCTGATGAAGATGAGCC 632
                 |||   |||   |||      :::   |||:::|||
     522 HisSerGluGlyGlnGlyGlyGlnGlyValGlnIleAspLysAsp..... 536

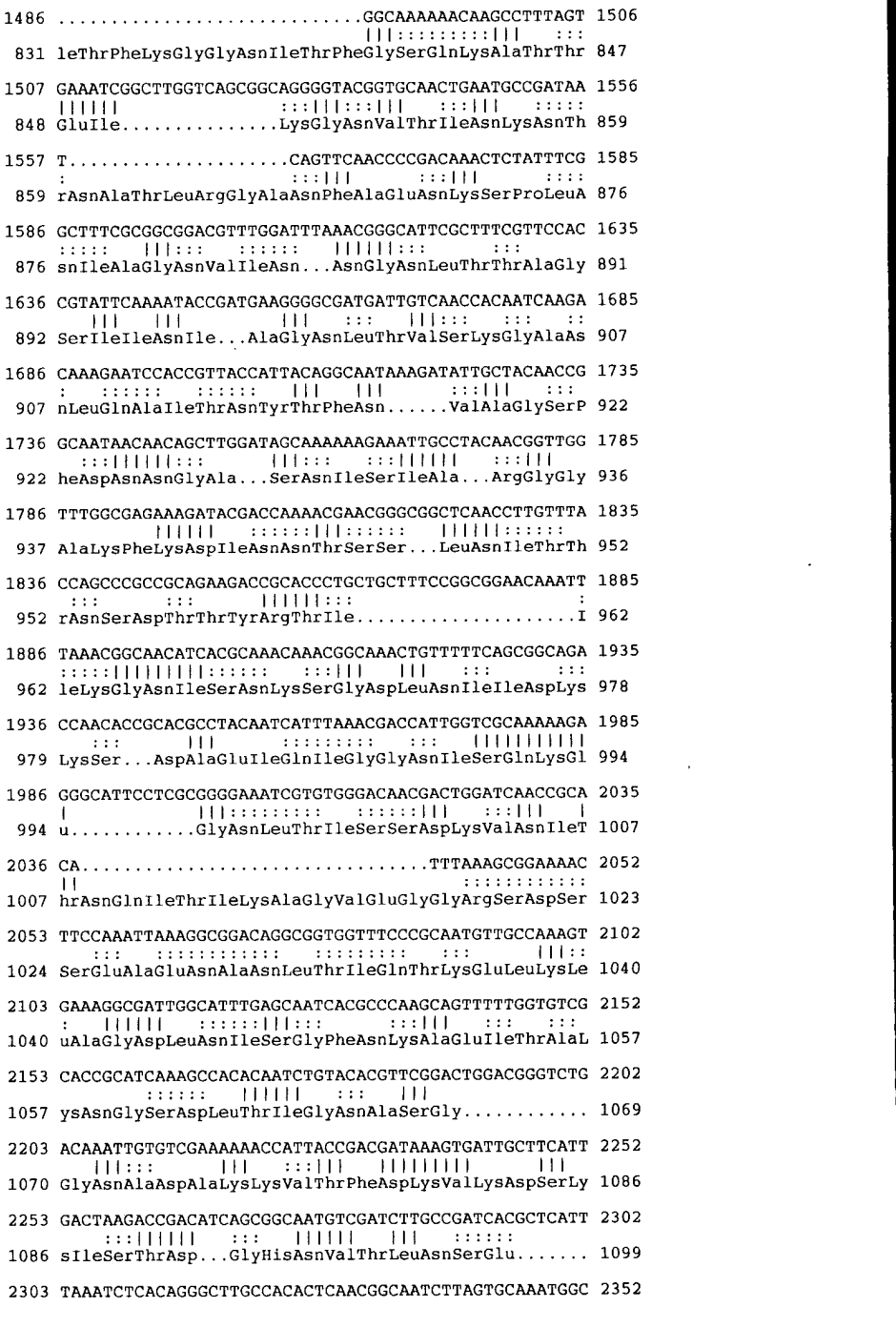
     633 CAATAACCGCGAAAGTTCATATCATATTGCAAGTGCGTATTCTTGGCTCG 682
                                |||:::|||
     537 .........................IleThrSerGlu............. 540

     683 TTGGTGGCAATACCTTTGCACAAAATGGATCAGGTGGT............ 720
          |||||||||                |||||||||
     541 ..GlyGlyAsnLeuThrIleTyr......SerGlyGlyTrpValAspVal 554

     721 ...GGCACAGTCAACTTAGGTAGTGAAAAAATTAAACATAGCCCATATGG 767
            :::::::::|||||||||                          ||
     555 HisLysAsnIleThrLeuGlySer.......................Gl 563

     768 TTTTTTACCAACAGGAGGCTCATTTGGCGAC.................A 799
         |||||||             :::   ||||||                |
     563 yPheLeuAsnIleThrThrLysGluGlyAspIleAlaPheGluAspLysS 580
```


800 GTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGGTTAATT 849
erGlyArgAsnAsnLeuThrIleThrAlaGln 590
850 AATGGGGTATTGCAAACGGGCAACCCCTATATAGGAAAAAGCAATGGCTT 899
...GlyThrIleThrSerGlyAsn SerAsnGlyPh 601
900 CCAGCTGGTTCGTAAAGATTGGTTCTATGATGAAATCTTTGCTGGAGATA 949
eArg PheAsnAsnValSerLeuAsnSerL 611
950 CCCATTCAGTATTCTACGAACCACGTCAAAATGGGAAATACTCTTTTAAC 999
eu GlyGlyLysLeuSerPheThr 618
1000 GACGATAAT 1008
AspSerArgGluAspArgGlyArgArgThrLysGlyAsnIleSerAsnLy 635
1009 AATGGCACAGGAAAAATCAATGCCAAAC 1036
sPheAspGlyThrLeuAsnIleSerGlyThrValAspIleSerMetLysA 652
1037 ATGAACACAATTCTCTGCCTAATAGATTAAAAACACGAACCGTT 1080
laProLysValSerTrpPheTyrArgAspLysGlyArgThrTyrTrpAsn 668
1081 CAATTGTTTAATGTTTCTTTATC 1103
ValThrThrLeuAsnValThrSerGlySerLysPheAsnLeuSerIleAs 685
1104 CGAGACAGCAAGAGAACCTGTTTATCATGCTGCAGGTGGTGTCAACAGTT 1153
pSerThr GlySerGlySerT 692
1154 ATCGACCCAGACTGAATAATGGAGAA AATATTTCCTTT 1191
hrGlyProSerIleArgAsnAlaGluLeuAsnGlyIleThrPheAsnLys 708
1192 ATTGACGAAGGAAAAGGCGAATTGATACTTACCAGCAA 1229
AlaThrPheAsnIleAlaGlnGlySer ThrAlaAs 720
1230 CATCAATCAAGGTGCTGGAGGATTATATTTCCAAGGAGATTTTACGGTCT 1279
nPheSerIleLysAlaSerIleMetProPheLysSerAsnAlaAsnTyrA 737
1280 CGCCTGAAAATAACGAAACTTGGCAAGGCGCGGGCGTTCATATCAGTGAA 1329
laLeuPheAsnGluAsp IleSerValSerGly 747
1330 GACAGTACCGTTACTTGGAAAGTAAACGGC 1359
GlyGlySerValAsnPheLysLeuAsnAlaSerSerSerAsnIleGlnTh 764
1360 GTGGCAAACGACCGCCTGTCCAAAATCGGCAAAGGCA 1396
rProGlyValIleIleLysSerGlnAsnPheAsnValSerGlyGlySerT 781
1397 CGCTGCACGTTCAAGCCAAAGGGGAAAACCAAGGCTCGATCAGCGTG 1443
hrLeuAsnLeuLysAlaGluGlySerThrGluThrAlaPheSerIleGlu 797
1444 GGCGACGGTACAGTCATTTTGGA 1466
AsnAspLeuAsnLeuAsnAlaThrGlyGlyAsnIleThrIleArgGlnVa 814
1467 TCAGCAGGCAGACGATAAA 1485
lGluGlyThrAspSerArgValAsnLysGlyValAlaAlaLysLysAsnI 831


1486 GGCAAAAAACAAGCCTTTAGT 1506
leThrPheLysGlyGlyAsnIleThrPheGlySerGlnLysAlaThrThr 847
1507 GAAATCGGCTTGGTCAGCGGCAGGGGTACGGTGCAACTGAATGCCGATAA 1556
GluIle LysGlyAsnValThrIleAsnLysAsnTh 859
1557 T CAGTTCAACCCCGACAAACTCTATTTCG 1585
rAsnAlaThrLeuArgGlyAlaAsnPheAlaGluAsnLysSerProLeuA 876
1586 GCTTTCGCGGCGGACGTTTGGATTTAAACGGGCATTCGCTTTCGTTCCAC 1635
snIleAlaGlyAsnValIleAsn AsnGlyAsnLeuThrThrAlaGly 891
1636 CGTATTCAAAATACCGATGAAGGGGCGATGATTGTCAACCACAATCAAGA 1685
SerIleIleAsnIle AlaGlyAsnLeuThrValSerLysGlyAlaAs 907
1686 CAAAGAATCCACCGTTACCATTACAGGCAATAAAGATATTGCTACAACCG 1735
nLeuGlnAlaIleThrAsnTyrThrPheAsn ValAlaGlySerP 922
1736 GCAATAACAACAGCTTGGATAGCAAAAAAGAAATTGCCTACAACGGTTGG 1785
heAspAsnAsnGlyAla SerAsnIleSerIleAla ArgGlyGly 936
1786 TTTGGCGAGAAAGATACGACCAAAACGAACGGGCGGCTCAACCTTGTTTA 1835
AlaLysPheLysAspIleAsnAsnThrSerSer LeuAsnIleThrTh 952
1836 CCAGCCCGCCGCAGAAGACCGCACCCTGCTGCTTTCCGGCGGAACAAATT 1885
rAsnSerAspThrThrTyrArgThrIle I 962
1886 TAAACGGCAACATCACGCAAACAAACGGCAAACTGTTTTTCAGCGGCAGA 1935
leLysGlyAsnIleSerAsnLysSerGlyAspLeuAsnIleIleAspLys 978
1936 CCAACACCGCACGCCTACAATCATTTAAACGACCATTGGTCGCAAAAAGA 1985
LysSer AspAlaGluIleGlnIleGlyGlyAsnIleSerGlnLysGl 994
1986 GGGCATTCCTCGCGGGGAAATCGTGTGGGACAACGACTGGATCAACCGCA 2035
u GlyAsnLeuThrIleSerSerAspLysValAsnIleT 1007
2036 CA TTTAAAGCGGAAAAC 2052
hrAsnGlnIleThrIleLysAlaGlyValGluGlyGlyArgSerAspSer 1023
2053 TTCCAAATTAAAGGCGGACAGGCGGTGGTTTCCCGCAATGTTGCCAAAGT 2102
SerGluAlaGluAsnAlaAsnLeuThrIleGlnThrLysGluLeuLysLe 1040
2103 GAAAGGCGATTGGCATTTGAGCAATCACGCCCAAGCAGTTTTTGGTGTCG 2152
uAlaGlyAspLeuAsnIleSerGlyPheAsnLysAlaGluIleThrAlaL 1057
2153 CACCGCATCAAAGCCACACAATCTGTACACGTTCGGACTGGACGGGTCTG 2202
ysAsnGlySerAspLeuThrIleGlyAsnAlaSerGly 1069
2203 ACAAATTGTGTCGAAAAAACCATTACCGACGATAAAGTGATTGCTTCATT 2252
GlyAsnAlaAspAlaLysLysValThrPheAspLysValLysAspSerLy 1086
2253 GACTAAGACCGACATCAGCGGCAATGTCGATCTTGCCGATCACGCTCATT 2302
sIleSerThrAsp GlyHisAsnValThrLeuAsnSerGlu 1099
2303 TAAATCTCACAGGGCTTGCCACACTCAACGGCAATCTTAGTGCAAATGGC 2352

```
                 :::   |||   ||||||:::   :::|||::::::
1100 .............ValLysThrSerAsnGlySerSerAsnAlaGlyAsn 1111

2353 GAT......ACACGTTATACAGTCAGC...CACAACGCCACCCAAAACGG 2393
     |||      |||     |||:::|||      :::::::|||   |||::
1112 AspAsnSerThrGlyLeuThrIleSerAlaLysAspValThrValAsnAs 1128

2394 CAACCTTAGCCTCGTGGGCAATGCCCAAGCAACATTTAATCAAGCCACAT 2443
     :|||::::::            ::::::   |||:::|||   ::::::
1128 nAsnValThr............SerHisLysThrIleAsnIleSerAlaA 1141

2444 TAAACGGCAACACATCGGCTTCGGGCAATGCTTCATTTAATCTAAGCGAC 2493
       ||||||:::::::::::   ::::::::::::|||   :::
1141 laAlaGlyAsnValThrThrLysGluGlyThrThrIleAsnAlaThrThr 1157

2494 CACGCCGTA............CAAAACGGCAGTCTGACGCTTTCCGGCAA 2531
        :::|||          |||||||||      |||::::::|||||
1158 GlySerValGluValThrAlaGlnAsnGly......ThrIleLysGlyAs 1172

2532 C...GCTAAGGCAAACGTAAGCCATTCCGCACTCAACGGTAATGTCTCCC 2578
     |   ::::::   ||||||:::   :::|||   :::   |||:::
1172 nIleThrSerGlnAsnValThrValThrAlaThrGlu...AsnLeuValT 1188

2579 TAGCCGATAAGGCAGTATTCCATTTTGAAAGC..............AGC 2613
       :::::::::||||||::::::      |||              |||
1188 hrThrGluAsnAlaValIleAsnAlaThrSerGlyThrValAsnIleSer 1204

2614 CGCTTTACCGGACAAATCAGCGGCGGCAAGGATACGGCA.........TT 2654
           ||||||:::|||:::||||||   :::::::::        ::
1205 ThrLysThrGlyAspIleLysGlyGlyIleGluSerThrSerGlyAsnVa 1221

2655 ACACTTAAAAGACAGCGAATGGACGCTGCCGTCAGGCACGGAATTAGGCA 2704
     :::::::      |||      ||||||            :::::::::|
1221 lAsnIleThrAlaSerGlyAsnThrLeu............LysValSerA 1234

2705 ATTTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCGCCAC 2754
     ||::::::   :::::::::|||:::|||   :::
1234 snIleThrGlyGlnAspValThrValThrAlaAsp.............. 1245

2755 GATGCGGCAGGGGCGCAAACCGGCAGTGCGACAGATGCGCCGCGCCGCCG 2804
           |||||||||   |||   :::|||
1246 ......AlaGlyAlaLeuThrThrThrAla..................... 1253

2805 TTCGCGCCGTTCGCGCCGTTCCCTATTATCCGTTACACCGCCAACTTCGG 2854
                     |||   :::|||:::|||      ::::::
1254 ................GlySerThrIleSerAlaThrThrGlyAsnAla. 1264

2855 TAGAATCCCGTTTCAACACGCTGACGGTAAACGGCAAATTGAACGGTCAG 2904
                |||      |||   :::|||   :::|||||||:::
1265 .............AsnIleThrThrLysThrGlyAspIleAsnGlyLys 1276

2905 GGAACATTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGACAAATT 2954

1276 ..................................................... 1276

2955 GAAGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCGGTCAACAAT. 3003
           :::|||||||||:::|||:::   ||||||:::::::::::::
1277 .......ValGluSerSerSerGlySerValThrLeuValAlaThrGlyA 1291

3004 ...........ACCGGCAACGAACCTGCAAGCCTCGAACAATTGACGGTA 3042
                 :::||||||      ::::::          :::|||:::
1291 laThrLeuAlaValGlyAsnIleSerGlyAsnThrValThrIleThrAla 1307

3043 GTGGAAGGAAAAGACAACAAACCGCTGTCCGAAAACCTTAAT........ 3084
       :::||||||   ::::::   ::::::::::::::|||
1308 AspSerGlyLysLeuThrSerThrValGlySerThrIleAsnGlyThrAs 1324

3085 .....................................TTCACCCTGCAAA 3097
                                          |||:::::::
1324 nSerValThrThrSerSerGlnSerGlyAspIleGluGlyThrIleSerG 1341

3098 ACGAACACGTCGATGCCGGCGCGTGGCGTTACCAACTCATCCGCAAAGAC 3147
     :::::   |||::::::   |||                    :::
```

```
1341 lyAsnThrValAsnValThrAla.....................SerThr 1350

3148 GGCGAGTTCCGCCTGCATAATCCGGTCAAAGAACAAGAGCTTTCCGACAA 3197
     |||:::::::   :::   |||   :::|||          ::::::   ||
1351 GlyAspLeuThrIleGlyAsnSerAlaLys.........ValGluAlaLy 1364

3198 ACTCGGCAAGGCAGAAGCCAAAAAACAGGCGGAAAAAGACAACGCGCAAA 3247
     |   |||   |||             ::::::   |||   :::::::|||:
1364 sAsnGlyAlaAlaThrLeuThrAlaGluSerGlyLysLeuThrThrGlnT 1381

3248 GCCTTGACGCGCTGATTGCGGCCGGGCGCGATGCCGTCGAAAAGACAGAA 3297
     ::   :::::::::   :::::::::                  :::|||
1381 hrGlySerSerIleThrSerSerAsn............GlyGlnThrThr 1393

3298 AGCGTTGCCGAACCGGCCCGGCAGGCAGGCGGGGAAAAT........... 3336
        :::|||:::   :::      ||||||:::   |||
1394 LeuThrAlaLysAspSerSerIleAlaGlyAsnIleAsnAlaAlaAsnVa 1410

3337 .............GTCGGCATTATGCAGGCGGAGGAAGAGAAAAAACGGG 3373
                 :::|||   :::   :::      :::::|||   :
1410 lThrLeuAsnThrThrGlyThrLeuThrThrThrGlyAspSerLys...I 1426

3374 TGCAGGCGGATAAAGACACCGCCTTGGCGAAACAGCGCGAAGCGGAAACC 3423
     :::::|||   :::   |||          :::   ::::::|||:::
1426 leAsnAlaThrSerGlyThrLeuThrIleAsnAlaLysAspAlaLysLeu 1442

3424 CGGCCGGCTACCACCGCCTTCCCCCGCGCCCGCCGCGCCCGCCGGGATTT 3473
           |||::::::::::
1443 AspGlyAlaAlaSerGly.............................. 1448

3474 GCCGCAACTGCAACCCCAACCGCAGCCCCAACCGCAGCGCGACCTGATCA 3523
                                          :::|||   :::::::
1449 ................................AspArgThrValValA 1454

3524 GCCGTTATGCCAATAGCGGTTTGAGTGAATTTTCCGCCACGCTCAACAGC 3573
     ::            ||||||   ::::::   :::|||      :::|||
1454 snAlaThrAsnAlaSerGlySerGlyAsnValThrAlaLysThrSerSer 1470

3574 GTTTTCGCCGTACAGGACGAATTAGAC..............CGCGTATT 3608
               :::   :::|||:::                 ::::::::
1471 SerValAsnIleThrGlyAspLeuAsnThrIleAsnGlyLeuAsnIleIl 1487

3609 TGCCGAAGACCGCCGCAACGCCGTTTGGACAAGCGGCATCCGG...GACA 3655
     :::|||:::   ||||||:::|||          |||   :::   |||:
1487 eSerGluAsnGlyArgAsnThrValArgLeuArgGlyLysGluIleAspV 1504

3656 CCAAACACTACCGTTCGCAAGATTTCCGCGCCTACCGCCAACAAACCGAC 3705
     ::|||:::          |||    ::::::   ::::::::::::
1504 alLysTyrIle......GlnPro.GlyValAlaSerValGluGluValIl 1518

3706 CTGCGCCAAATCGGTATGCAGAAAAACC..................... 3733
          ||||||   |||   :::|||:::
1518 eGluAlaLysArgValLeuGluLysValLysAspLeuSerAspGluGluA 1535

3734 .....TCGGCAGCGGGCGCGTCGGCATCCTGTTTTCGCACAACCGGACCG 3778
           :::   |||      :::::::|||:::   |||   :::   |||
1535 rgGluThrLeuAlaLysLeuGlyValSerAlaValArgPheValGluPro 1551

3779 AAAACACCTTCGACGACGGCATCGGCAACTCGGCACGGCTTGCCCACGGC 3828
     ::::::      |||:::          |||:::::::   :::   |||
1552 AsnAsnAlaIleThrVal......AsnThrGlnAsnGluPheThrThrLy 1566

3829 GCCGTTTTCGGGCAATACGGCATCGACAGGTTCTACA......TCGGCAT 3872
      |||   |||       |||   |||   |||::::::       |||::::
1566 sProSerSerGlnValThrIleSerGluGlyLysAlaCysPheSerSerG 1583

3873 CAGCGCGGGCGCGGGTTTTAGCAGCGGCAGCC 3904
     ::   .   |||||||||   :::   :::|||
1583 lyAsnGlyAlaArgValCysThrAsnValAla 1593

seq_name: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-038-682-4

seq_documentation_block:
```

```
; Sequence 4, Application US/08038682
; Patent No. 5549897
;  GENERAL INFORMATION:
;    APPLICANT:  BARENKAMP, STEPHEN J
;    APPLICANT:  ST. GEME III, JOSEPH W
;    TITLE OF INVENTION:  HIGH MOLECULAR WEIGHT SURFACE PROTEINS
;    TITLE OF INVENTION:  OF NON-TYPEABLE HAEMOPHILUS
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Shoemaker and Mattare, Ltd
;      STREET:  2001 Jefferson Davis Hwy., 1203 Crystal Plaza
;      STREET:  Bldg. 1
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202-0286
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/038,682
;      FILING DATE:  16-MAR-1993
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BERKSTRESSER, JERRY W
;      REGISTRATION NUMBER:  22,651
;      REFERENCE/DOCKET NUMBER:  1038-293
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 415-0810
;      TELEFAX:  (703) 415-0813
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1477 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-038-682-4


alignment_scores:
            Quality:  282.50              Length:   1388
              Ratio:  0.428                Gaps:     61
 Percent Similarity:  47.550   Percent Identity: 19.380

alignment_block:
 US-09-303-518D-649 x US-08-038-682-4   ..

 Align seg 1/1  to: US-08-038-682-4  from: 1  to: 1477

     175 GAAAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAA 224
         :::|||:::|||   :::::::|||:::::::   :::|||::::::
     219 LysAsnGluGlyValIleSerValAsnGlyGlySerIleSerLeu..... 233

     225 CAAAAAAGGGGAGTTGGTCGGCAAATCAATGACAAAAGCC.......... 264
                    |||:::|||::::::::::|||   :::
     234 .............LeuAlaGlyGlnLysIleThrIleSerAspIleIleA 246

     265 ..CCGATGATTGATTTTTCTGTGGTGTCGCGTAACGGCGTGGCGGCATTG 312
           |||   |||   :::|||::::::::::   ::::::   |||:::
     246 snProThrIleThrTyrSerIleAlaAlaProGluAsnGluAlaValAsn 262

     313 GTGGGCGATCAATATATTGTGAGCGTGGCACATAACGGCGGCTATAACAA 362
         :::||||||               :::      :::|||||||      ||
     263 LeuGlyAsp..............IlePheAlaLysGlyGlyAsnIleAs 274

     363 CGTT..........................GATTTTGGTGCGGAAGGAA 385
         ||||                          :::::::|||::::::
     274 nValArgAlaAlaThrIleArgAsnGlnGlyLysLeuSerAlaAspServ 291

     386 GAAATCCCGATCAACATCGTTTTACTTATAAAATTGTGAAACGGAATAAT 435

         :::   |||:::                        :::|||::::::::::
     291 alSerLysAspLysSerGlyAsnIleValLeuSerAlaLysGluGlyGlu 307

     436 TATAAAGCAGGG..............ACTAAAGGCCATCCTTATGGCGG 470
           :::   |||                  :::   :::      |||||
     308 AlaGluIleGlyGlyValIleSerAlaGlnAsnGlnGlnAlaLysGlyGl 324

     471 CGATTATCATATGCCGCGTTTGCATAAATTTGTCACAGATGCAGAACCTG 520
         |         :::                  :::      :::      :
     324 yLysLeuMetIleThrGlyAspLysValThrLeuLysThrGlyAlaValI 341

     521 TTGAAATGACCAGTTATATGGATGGGCGGAAATATATCGATCAAAATAAT 570
         ::::::::::::::         |||:::   |||:::      ::::::
     341 leAspLeuSerGlyLysGluGlyGlyGluThrTyrLeuGlyGlyAspGlu 357

     571 TACCCTGACCGTGTTCGTATTGGGGCAGGCAGGCAATATTGGCGATCTGA 620
                               |||   |||::::::      :::
     358 Arg.................GlyGluGlyLysAsnGlyIleGlnLeuAl 368

     621 TGAAGATGAGCCCAATAACCGCGAAAGTTCATATCATATTGCAAGT.... 666
         :::           ::::::   |||:::   ::::::::::::
     368 aLysLysThrSerLeuGluLysGlySerThrIleAsnValSerGlyLysG 385

     667 .............GCGTATTCTTGG............CTCGTTGGTGGC 690
                     |||      |||            |||:::   |||
     385 luLysGlyGlyPheAlaIleValTrpGlyAspIleAlaLeuIleAspGly 401

     691 AATACCTTTGCACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAG 740
         |||      ||||||   |||||||||
     402 AsnIleAsnAlaGln...GlySerGly..................... 409

     741 TGAAAAAATTAAACATAGCCCATATGGTTTTTTACCAACAGGAGGCTCAT 790
                                  :::      |||||||||
     410 .......................AspIleAlaLysThrGlyGlyPheV 418

     791 TTGGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAG 840
               ||||||      :::||||||   |||
     418 alGluThrSerGlyHisAspLeuPheIleLysAsp.............. 429

     841 TGGTTAATTAATGGGGTATTGCAAACGGGCAACCCCTATATAGGAAAAAG 890
               |||:::::::::::::::
     430 .........AsnAlaIleValAspAla..................... 435

     891 CAATGGCTTCCAGCTGGTTCGTAAAGATTGGTTCTATGATGAAATCTTTG 940
                                  |||:::|||:::   |||
     436 .....................LysGluTrpLeuLeuAsp......... 441

     941 CTGGAGATACCCATTCAGTATTCTACGAACCACGTCAAAATGGGAAATAC 990
           |||:::   |||:::      |||              ::::::
     442 ..PheAspAsnValSerIleAsnAlaGluAspProLeuPheAsnAsnThr 457

     991 TCTTTTAACGACGAT...AATAATGGCACAGGAAAAATCAATGCCAAACA 1037
         ::::::||||||:::      :::||||||||||:::   :::
     458 GlyIleAsnAspGluPheProThrGlyThrGlyGluAlaSerAspProLy 474

    1038 TGAACACAATTCTCTGCCTAATAGATTAAAAACACGAACCGTTCAATTGT 1087
           :::::::::::|||   :::   |||   :::   |||::::::   :
     474 sLysAsnSerGluLeuLysThrThrLeuThrAsnThrThrIleSerAsnT 491

    1088 TT.....................AATGTTTCTTTATCCGAGACAGCAAGA 1116
         ::                        |||:::::::   |||:::
     491 yrLeuLysAsnAlaTrpThrMetAsnIleThrAlaSerArgLysLeuThr 507

    1117 GAACCTGTTTATCATGCTGCAGGTGGTGTCAACAGTTATCGACCCAGACT 1166
                                     ||||||||             ::
     508 ..........................ValAsnSer.........SerIl 512

    1167 GAATAATGGAGAAAATATTTCCTTTATTGACGAAGGAAAAGGCGAATTGA 1216
         :|||   |||:::|||      :::|||   ::::::||||||
     512 eAsnIleGlySerAsnSerHisLeuIleLeuHisSerLysGly..... 526

    1217 TACTTACCAGCAACATCAATCAAGGTGCTGGAGGATTATATTTCCAAGGA 1266
                              |||   :::|||||||:::   ::::::|||
```

```
 527 ...................GlnArgGlyGlyGlyValGlnIleAspGly 536

1267 GATTTTACGGTCTCGCCTGAAAATAACGAAACTTGGCAAGGCGCGGGCGT 1316
     |||:::|||   :::      |||         :::::::|||:::   ||
 537 AspIleThrSerLysGlyGlyAsnLeuThrIleTyrSerGlyGlyTrpVa 553

1317 TCATATCAGTGAAGACAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAA 1366
     |   :::   ::::::   |||:::
 553 lAspValHisLysAsnIleThrLeu..................... 561

1367 ACGACCGCCTGTCCAAAATCGGCAAAGGCACGCTGCACGTTCAAGCCAAA 1416

 561 ................................................. 561

1417 GGGGAAAACCAAGGCTCGATCAGCGTGGGCGACGGTACAGTCATTTTGGA 1466
        ::|||||   ::::::::      ::::::|||   :::
 562 ......AspGlnGlyPheLeuAsnIleThrAlaAlaSerValAlaPhe.. 575

1467 TCAGCAGGCAGACGATAAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCT 1516
      :::   :::::::::|||:::::      |||   :::::|||
 576 .GluGlyGlyAsnAsnLysAlaArgAspAlaAlaAsnAlaLysIle.... 590

1517 TGGTCAGCGGCAGGGGTACGGTGCAACTGAATGCCGATAATCAGTTCAAC 1566
           ::::::|||||||||   ::::::::::::::::::
 591 .....ValAlaGlnGlyThrValThrIleThrGlyGluGlyLys...... 603

1567 CCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAACGG 1616
                    ||||||:::::::::::::::|||||||
 604 .................AspPheArgAlaAsnAsnValSerLeuAsnGl 614

1617 GCATTCGCTTTCGTTCCACCGTATTCAAAATACCGATGAAGGGGCGATGA 1666
     |                          |||   :::|||      |
 614 y............................ThrGlyLysGlyLeuAsnI 621

1667 TTGTC............AACCACAATCAAGACAAAGAATCCACCGTTACC 1704
     ||:::          |||   :::::::::   ::::::|||::::::
 621 leIleSerSerValAsnAsnLeuThrHisAsnLeuSerGlyThrIleAsn 637

1705 ATTACAGGCAATAAAGATATTGCTACAACCGGCAATAACAACAGCTTGGA 1754
     |||:::||||||      |||      |||
 638 IleSerGlyAsnIleThrIleAsnGlnThr..................... 647

1755 TAGCAAAAAAGAAATTGCCTACAACGGTTGGTTTGGCGAGAAAGATACG. 1803
      ::::::|||:::   :::|||      |||      :::   |||:::
 648 .ThrArgLysAsnThrSerTyr......TrpGlnThrSerHisAspSerH 662

1804 ......................ACCAAAACGAACGGGCGGCTCAACCTT 1830
                            :::|||:::::::::::::::::::
 662 isTrpAsnValSerAlaLeuAsnLeuGluThrGlyAlaAsnPheThrPhe 678

1831 GTTTACCAGCCCGCCGCAGAAGACCGCACCCTGCTGCTT........... 1869
     :::          :::::::::::::::   |||
 679 IleLysTyrIleSerSerAsnSerLysGlyLeuThrThrGlnTyrArgSe 695

1870 .TCCGGCGGAACAAATTTAAACGGCAACATCACGCAAACAAACGGCAAAC 1918
      |||:::|||:::|||:::||||||           :::||||||::::
 695 rSerAlaGlyValAsnPheAsnGly............ValAsnGlyAsnM 708

1919 TGTTTTTCAGCGGCAGACCAACACCGCACGCCTACAATCATTTAAACGAC 1968
     ::   |||
 708 etSerPhe..................................... 710

1969 CATTGGTCGCAAAAAGAGGGCATTCCTCGCGGGGAAATCGTGTGGGACAA 2018
           :::   |||||||||         :::      :::      ||
 711 ......AsnLeuLysGluGlyAlaLysValAsnPheLysLeuLysProAs 725

2019 CGACTGGATCAACCGCACA.................TTTAAAGCGGAAA 2050
     |:::   :::|||   :::                  |||   |||   |
 725 nGluAsnMetAsnThrSerLysProLeuProIleArgPheLeuAla...A 741

2051 ACTTCCAAATTAAAGGCGGACAGGCGGTGGTT................. 2082
     ||:::         ||||||   :::|||
 741 snIleThrAlaThrGlyGlyGlySerValPhePheAspIleTyrAlaAsn 757

2083 ...TCCCGCAATGTTGCCAAAGTGAAA...GGCGATTGGCATTTGAGCAA 2126
        |||   :::   |||:::::::|||   ::::::   ::::::|||||
  758 HisSerGlyArgGlyAlaGluLeuLysMetSerGluIleAsnIleSerAs 774

2127 TCACGCCCAAGCAGTTTTTGGTGTCGCACCGCATCAAAGCCACACAATCT 2176
     |   |||:::   ::::::                 :::|||||||
  774 nGlyAlaAsnPheThrLeu..............AsnSerHis....... 783

2177 GTACACGTTCGGACTGGACGGGTCTGACAAATTGTGTCGAAAAAACCATT 2226
       :::|||:::|||   :::         |||      :::   ||||||
  784 ..ValArgGlyAspAspAlaPheLysIleAsn...LysAspLeuThrIle 798

2227 ACCGACGATAAAGTGATTGCTTCATTGACTAAGACCGACATCAGCGGCAA 2276
     :::      :::         ||||||   :::|||      :::
  799 AsnAlaThrAsnSerAsnPheSerLeuArgGlnThrLysAspAspPheTy 815

2277 TGTCGATCTTGCCGATCACGCTCATTTAAATCTCACAGGGCTTGCCACAC 2326
            |||   :::|||      :::      :::::::::   |
  815 rAspGlyTyrAlaArgAsnAlaIleAsnSerThrTyrAsnIleSerIleL 832

2327 TCAACGGCAATCTTAGTGCAAATGGCGATACACGTTATACAGTCAGCCAC 2376
     ||:::|||||:::::   :::|||
  832 euGlyGlyAsnValThrLeuGlyGly..................... 840

2377 AACGCCACCCAAAACGGCAACCTTAGCCTCGTGGGCAATGCCCAAGCAAC 2426
            ||||||::::::   |||::::::||||||          ||
  841 .........GlnAsnSerSerSerSerIleThrGlyAsn......IleTh 852

2427 ATTTAATCAAGCC.........ACATTAAACGGCAACACATCGGCTTCGG 2467
     |::::::::::|||         ||||||:::::::|||::::::   :::
  852 rIleGluLysAlaAlaAsnValThrLeuGluAlaAsnAsnAlaProAsnG 869

2468 GCAATGCTTCATTTAATCTAAGCGACCACGCCGTACAAAACGGCAGTCTG 2517
       :::         |||:::   |||:::::::::::   |||||||||
  869 lnGln.........AsnIleArgAspArgValIleLysLeuGlySerLeu 882

2518 ACGCTTTCCGGCAACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACGG 2567
        :::::::|||:::   :::   :::   :::::::::   :::::::||
  883 LeuValAsnGlySerLeuSerLeuThrGlyGluAsnAlaAspIleLysGl 899

2568 TAATGTCTCCCTAGCCGATAAGGCAGTATTCCAT.........TTTGAAA 2608
     ||||:::::::::::::::::::|||:::|||                ::::
  899 yAsnLeuThrIleSerGluSerAlaThrPheLysGlyLysThrArgAspT 916

2609 GCAGCCGCTTTACCGGACAAATCAGCGGCGGCAAGGATACGGCATTACAC 2658
     ::   ::::::||||||:::::::::::::::      :::   ::::::
  916 hrLeuAsnIleThrGlyAsnPheThrAsnAsnGlyThrAlaGluIleAsn 932

2659 TTAAAAGACAGCGAATGGACGCTGCCGTCAGGCACGGAATTAGGCAATTT 2708
     :::   ::::::         |||   :::   |||:::   |||:::||
  933 IleThrGlnGlyValValLysLeuGlyAsnValThrAsnAspGlyAspLe 949

2709 AAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCGCCACGATG 2758
     ||||:::   :::
  949 uAsnIleThrThr................................... 953

2759 CGGCAGGGGCGCAAACCGGCAGTGCGACAGATGCGCCGCGCCGCCGTTCG 2808
                                                 ::::::
  954 .....................................HisAla 955

2809 CGCCGTTCGCGCCGTTCCCTATTATCCGT....TACACCGCCAACTTCGG 2854
     :::|||::::::||||||:::::::::                 :::   ||
  956 LysArgAsnGlnArgSerIleIleGlyGlyAspIleIleAsnLysLysGl 972

2855 TAGAATCCCGTTTCAACACGCTGACGGTAAACGGCAAATTGAACGGTCAG 2904
     |   :::   :::      :::::::::      :::|||:::   :::|
  972 ySerLeuAsnIleThrAspSerAsnAsnAspAlaGluIleGlnIleGlyG 989

2905 GGAACATTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGACAAATT 2954
     ||||||||            |||:::|||   :::    |||||||||::
  989 lyAsnIleSerGlnLysGluGlyAsnLeuThrIle.SerSerAspLysIl 1005
```

```
2955 GAAGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCGGTCAACAATA 3004
     ::::::::::
1005 eAsnIleThr..................................... 1008

3005 CCGGCAACGAACCTGCAAGCCTCGAACAATTGACGGTAGTGGAAGGAAAA 3054
                          :::|||:::|||:::   :::|||
1009 ......................LysGlnIleThrIleLysLysGlyIle 1017

3055 GACAACAAACCGCTGTCC............GAAAACCTTAATTTCACCCT 3092
     |||::::::      |||           :::|||   |||:::|||::
1018 AspGlyGluAspSerSerSerAspAlaThrSerAsnAlaAsnLeuThrIl 1034

3093 GCAAAACGAACACGTCGAT.....................GCCGGCGCGT 3121
     ::::::::::::::::                        :::|||
1034 eLysThrLysGluLeuLysLeuThrGluAspLeuSerIleSerGlyPheA 1051

3122 GGCGTTACCAACTCATCCGCAAAGACGGC...GAGTTCCGCCTGCATAAT 3168
       :::   ::::::      |||||||||   ::::::   :::   |||
1051 snLysAlaGluIleThrAlaLysAspGlyArgAspLeuThrIleGlyAsn 1067

3169 CCGGTCAAAGAACAAGAGCTTTCCGACAAACTCGGCAAGGCAGAAGCCAA 3218
                        :::|||      :::   |||||||||
1068 Ser.................AsnAspGlyAsnSerGlyAlaGluAla.. 1077

3219 AAAACAGGCGGAAAAAGACAACGCGCAAAGCCTTGACGCGCTGATTGCGG 3268
      |||   :::      :::|||::::::      |||:::   |||:::|
1078 .LysThrValThrPheAsnAsnValLys......AspSerLysIleSerA 1092

3269 CCGGGCGCGATGCCGTCGAAAAGACAGAAAGCGTTGCCGAACCGGCCCGG 3318
     ||            |||      ::::::::::|||         :::
1092 laAspGlyHisAsnValThrLeuAsnSerLysValLysThrSerSerSer 1108

3319 CAGGCAGGCGGGGAAAATGTCGGCATTATGCAGGCGGAGGAAGAGAAAAA 3368
     :::::::|||   |||:::
1109 AsnGlyGlyArgGluSer.............................. 1114

3369 ACGGGTGCAGGCGGATAAAGACACCGCCTTGGCGAAACAGCGCGAAGCGG 3418
            ::::::|||:::||||||:::|||
1115 .......AsnSerAspAsnAspThrGlyLeu................... 1122

3419 AAACCCGGCCGGCTACCACCGCCTTCCCCCGCGCCCGCCGCGCCCGCCGG 3468
                     |||         :::      ::::::::::::
1123 .................ThrIleThrAlaLysAsnValGluValAsnLys 1133

3469 GATTTGCCGCAACTGCAACCCCAACCGCAGCCCCAACCGCAGCGCGACCT 3518
     |||:::   :::|||
1134 AspIleThrSerLeu................................... 1138

3519 GATCAGCCGTTATGCCAATAGCGGTTTGAGTGAATTTTCCGCCACGCTCA 3568
            :::   :::|||         ||||||      :::||| :::
1139 .......LysThrValAsnIleThrAlaSerGluLysValThrThr.Thr 1152

3569 ACAGCGTTTTCGCCGTACAGGACGAATTAGACCGCGTATTTGCCGAAGAC 3618
     ::::::   :::         ||||||      |||   :::
1153 AlaGlySerThrIleAsnAlaThrAsnGlyLysAlaSerIleThrThrLy 1169

3619 CGCCGCAACGCCGTTTGGACAAGCGGCATCCGGGACACCAAACACTACCG 3668
      ::::::         |||      :::::::::|||   ::::::::::
1169 sThrGlyAspIleSerGlyThrIleSerGlyAsnThrValSerValSerA 1186

3669 TTCGC...AAGATTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCAAA 3715
     ::        :::::::::   :::::::::|||      :::   :::
1186 laThrValAspLeuThrThrLysSerGlySerLysIleGluAlaLysSer 1202

3716 TCGGTATGCAGAAAAACCTCGGCAGCGGGCGCGTCGGCATCCTGTTTTCG 3765
     :::   ::::::   |||||||||:::|||:::   ::::::
1203 GlyGluAlaAsnValThrSerAlaThrGlyThrIleGlyGly........ 1216

3766 CACAACCGGACCGAAAACACCTTCGACGACGGCATCGGCA.......... 3805
      |||   :::   :::|||   ::::::||||||:::|||
1217 .ThrIleSerGlyAsnThrValAsnValThrAlaAsnAlaGlyAspLeuT 1233

3806 ........................ACTCGGCACGGCTTGCCCACGGCG 3829
```

```
                                    :::   :::|||      |||
1233 hrValGlyAsnGlyAlaGluIleAsnAlaThrGluGlyAlaAlaThrLeu 1249

3830 CCGTTTTCGGGCAATACGGCATCGACA......GGTTCTACATCGGCATC 3873
            :::||||||||||   :::|||      |||||||:::   :::||
1250 ThrAlaThrGlyAsnThrLeuThrThrGluAlaGlySerSerIleThrSe 1266

3874 AGCGCGGGCGCGGGTTTTAGCAGCGGCAGCCTTTCAGACGGCATCGGAGG 3923
     |:::::::::::::|||          |||      |||:::::::|||   |
1266 rThrLysGlyGlnValAspLeuLeuAla......GlnAsnGlySerIleA 1281

3924 CAAAATCCGCCCGCCGCGTGCTGCATTACGGCATTCAGGCACGATACCGCG 3973
     ||   |||      |||:::   :::|||                 |||
1281 laGlySerIleAsnAlaAlaAsnValThrLeuAsnThrThrGlyThrLeu 1297

3974 CCGGTTTCGGCGGATTCGGCATCGAACCGCACATCGGCGCAACGCGCTAT 4023
        :::   |||   |||      :::   ||||||:::      :::||
1298 ThrThrValAlaGlySerAspIleLysAlaThrSerGlyThrLeuValIl 1314

4024 TTCGTCCAAA 4033
     |::::::|||
1314 eAsnAlaLys 1317
```

```
seq_name: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-302-832-4

seq_documentation_block:
; Sequence 4, Application US/08302832
; Patent No. 5603938
;  GENERAL INFORMATION:
;    APPLICANT:  Barenkamp, Stephen J
;    TITLE OF INVENTION:  High Molecular Weight Surface Proteins
;    TITLE OF INVENTION:  of No. 5603938-Typeable Haemophilus
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Shoemaker and Mattare, Ltd.
;      STREET:  2001 Jefferson Davis Hwy., 1203 Crystal Plaza
;      STREET:  Bldg. 1
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202-0286
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/302,832
;      FILING DATE:  16-SEP-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9205704.1
;      FILING DATE:  16-MAR-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US pct/us93/02166
;      FILING DATE:  16-MAR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Berkstresser, Jerry W
;      REGISTRATION NUMBER:  22,651
;      REFERENCE/DOCKET NUMBER:  1038-404
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 415-0810
;      TELEFAX:  (703) 415-0813
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1477 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-302-832-4


alignment_scores:
```

```
          Quality:  282.50          Length:   1388
            Ratio:  0.428             Gaps:     61
Percent Similarity:  47.550   Percent Identity: 19.380

alignment_block:
US-09-303-518D-649 x US-08-302-832-4   ..

Align seg 1/1  to: US-08-302-832-4  from: 1  to: 1477

    175 GAAAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAA 224
        :::|||:::|||   :::::::|||:::::   :::|||::::::
    219 LysAsnGluGlyValIleSerValAsnGlyGlySerIleSerLeu..... 233

    225 CAAAAAAGGGGAGTTGGTCGGCAAATCAATGACAAAAGCC.......... 264
                    |||:::|||:::::::::|||   :::
    234 .............LeuAlaGlyGlnLysIleThrIleSerAspIleIleA 246

    265 ..CCGATGATTGATTTTTCTGTGGTGTCGCGTAACGGCGTGGCGGCATTG 312
          |||   |||   :::|||::::::::::   ::::::   |||:::
    246 snProThrIleThrTyrSerIleAlaAlaProGluAsnGluAlaValAsn 262

    313 GTGGGCGATCAATATATTGTGAGCGTGGCACATAACGGCGGCTATAACAA 362
        :::||||||               :::      :::||||||      ||
    263 LeuGlyAsp..............IlePheAlaLysGlyGlyAsnIleAs 274

    363 CGTT..........................GATTTTGGTGCGGAAGGAA 385
        ||||                          ::::::|||::::::
    274 nValArgAlaAlaThrIleArgAsnGlnGlyLysLeuSerAlaAspServ 291

    386 GAAATCCCGATCAACATCGTTTTACTTATAAAATTGTGAAACGGAATAAT 435
          :::   |||:::                    :::|||::::::::
    291 alSerLysAspLysSerGlyAsnIleValLeuSerAlaLysGluGlyGlu 307

    436 TATAAAGCAGGG..............ACTAAAGGCCATCCTTATGGCGG 470
           :::   |||                :::   :::      |||||
    308 AlaGluIleGlyGlyValIleSerAlaGlnAsnGlnGlnAlaLysGlyGl 324

    471 CGATTATCATATGCCGCGTTTGCATAAATTTGTCACAGATGCAGAACCTG 520
        |         :::                :::      :::      :
    324 yLysLeuMetIleThrGlyAspLysValThrLeuLysThrGlyAlaValI 341

    521 TTGAAATGACCAGTTATATGGATGGGCGGAAATATATCGATCAAAATAAT 570
        ::::::::::::::         |||:::   |||:::      ::::::
    341 leAspLeuSerGlyLysGluGlyGlyGluThrTyrLeuGlyGlyAspGlu 357

    571 TACCCTGACCGTGTTCGTATTGGGGCAGGCAGGCAATATTGGCGATCTGA 620
                           |||   |||::::::      :::
    358 Arg.................GlyGluGlyLysAsnGlyIleGlnLeuAl 368

    621 TGAAGATGAGCCCAATAACCGCGAAAGTTCATATCATATTGCAAGT.... 666
        :::            ::::::   |||:::   ::::::::::::
    368 aLysLysThrSerLeuGluLysGlySerThrIleAsnValSerGlyLysG 385

    667 .............GCGTATTCTTGG............CTCGTTGGTGGC 690
                    |||      |||           |||:::   |||
    385 luLysGlyGlyPheAlaIleValTrpGlyAspIleAlaLeuIleAspGly 401

    691 AATACCTTTGCACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAG 740
        |||      ||||||   |||||||||
    402 AsnIleAsnAlaGln...GlySerGly..................... 409

    741 TGAAAAAATTAAACATAGCCCATATGGTTTTTTACCAACAGGAGGCTCAT 790
                                 ::      |||||||||
    410 ........................AspIleAlaLysThrGlyGlyPheV 418

    791 TTGGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAG 840
              ||||||      :::||||||   |||
    418 alGluThrSerGlyHisAspLeuPheIleLysAsp............... 429

    841 TGGTTAATTAATGGGGTATTGCAAACGGGCAACCCCTATATAGGAAAAAG 890
              |||:::::::::::::::
    430 .........AsnAlaIleValAspAla...................... 435

    891 CAATGGCTTCCAGCTGGTTCGTAAAGATTGGTTCTATGATGAAATCTTTG 940
                                 |||:::|||:::   |||
    436 .....................LysGluTrpLeuLeuAsp......... 441

    941 CTGGAGATACCCATTCAGTATTCTACGAACCACGTCAAAATGGGAAATAC 990
            |||:::   |||:::      |||             ::::::
    442 ..PheAspAsnValSerIleAsnAlaGluAspProLeuPheAsnAsnThr 457

    991 TCTTTTAACGACGAT...AATAATGGCACAGGAAAAATCAATGCCAAACA 1037
        ::::::||||||:::      :::||||||||||:::   :::
    458 GlyIleAsnAspGluPheProThrGlyThrGlyGluAlaSerAspProLy 474

   1038 TGAACACAATTCTCTGCCTAATAGATTAAAAACACGAACCGTTCAATTGT 1087
         ::::::::::::|||   :::   |||   :::   |||::::::   :
    474 sLysAsnSerGluLeuLysThrThrLeuThrAsnThrThrIleSerAsnT 491

   1088 TT.....................AATGTTTCTTTATCCGAGACAGCAAGA 1116
        ::                     |||::::::   |||:::
    491 yrLeuLysAsnAlaTrpThrMetAsnIleThrAlaSerArgLysLeuThr 507

   1117 GAACCTGTTTATCATGCTGCAGGTGGTGTCAACAGTTATCGACCCAGACT 1166
                                |||||||||           ::
    508 ..........................ValAsnSer.........SerIl 512

   1167 GAATAATGGAGAAAATATTTCCTTTATTGACGAAGGAAAAGGCGAATTGA 1216
        :|||   |||:::|||      :::|||   ::::::||||||
    512 eAsnIleGlySerAsnSerHisLeuIleLeuHisSerLysGly....... 526

   1217 TACTTACCAGCAACATCAATCAAGGTGCTGGAGGATTATATTTCCAAGGA 1266
                              |||   :::||||||:::   ::::::|||
    527 ...................GlnArgGlyGlyGlyValGlnIleAspGly 536

   1267 GATTTTACGGTCTCGCCTGAAAATAACGAAACTTGGCAAGGCGCGGGCGT 1316
        |||:::|||   :::      |||          ::::::|||:::   ||
    537 AspIleThrSerLysGlyGlyAsnLeuThrIleTyrSerGlyGlyTrpVa 553

   1317 TCATATCAGTGAAGACAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAA 1366
        |   :::   ::::::   |||:::
    553 lAspValHisLysAsnIleThrLeu........................ 561

   1367 ACGACCGCCTGTCCAAAATCGGCAAAGGCACGCTGCACGTTCAAGCCAAA 1416

    561 .................................................. 561

   1417 GGGGAAAACCAAGGCTCGATCAGCGTGGGCGACGGTACAGTCATTTTGGA 1466
              :::|||||||   :::::::::      ::::::|||   :::
    562 ......AspGlnGlyPheLeuAsnIleThrAlaAlaSerValAlaPhe.. 575

   1467 TCAGCAGGCAGACGATAAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCT 1516
         :::   :::::::::|||::::::      |||   ::::::|||
    576 .GluGlyGlyAsnAsnLysAlaArgAspAlaAlaAsnAlaLysIle.... 590

   1517 TGGTCAGCGGCAGGGGTACGGTGCAACTGAATGCCGATAATCAGTTCAAC 1566
                ::::::||||||||   ::::::::::::::::::
    591 .....ValAlaGlnGlyThrValThrIleThrGlyGluGlyLys...... 603

   1567 CCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAACGG 1616
                              ||||||:::::::::::::::|||||||
    604 ..................AspPheArgAlaAsnAsnValSerLeuAsnGl 614

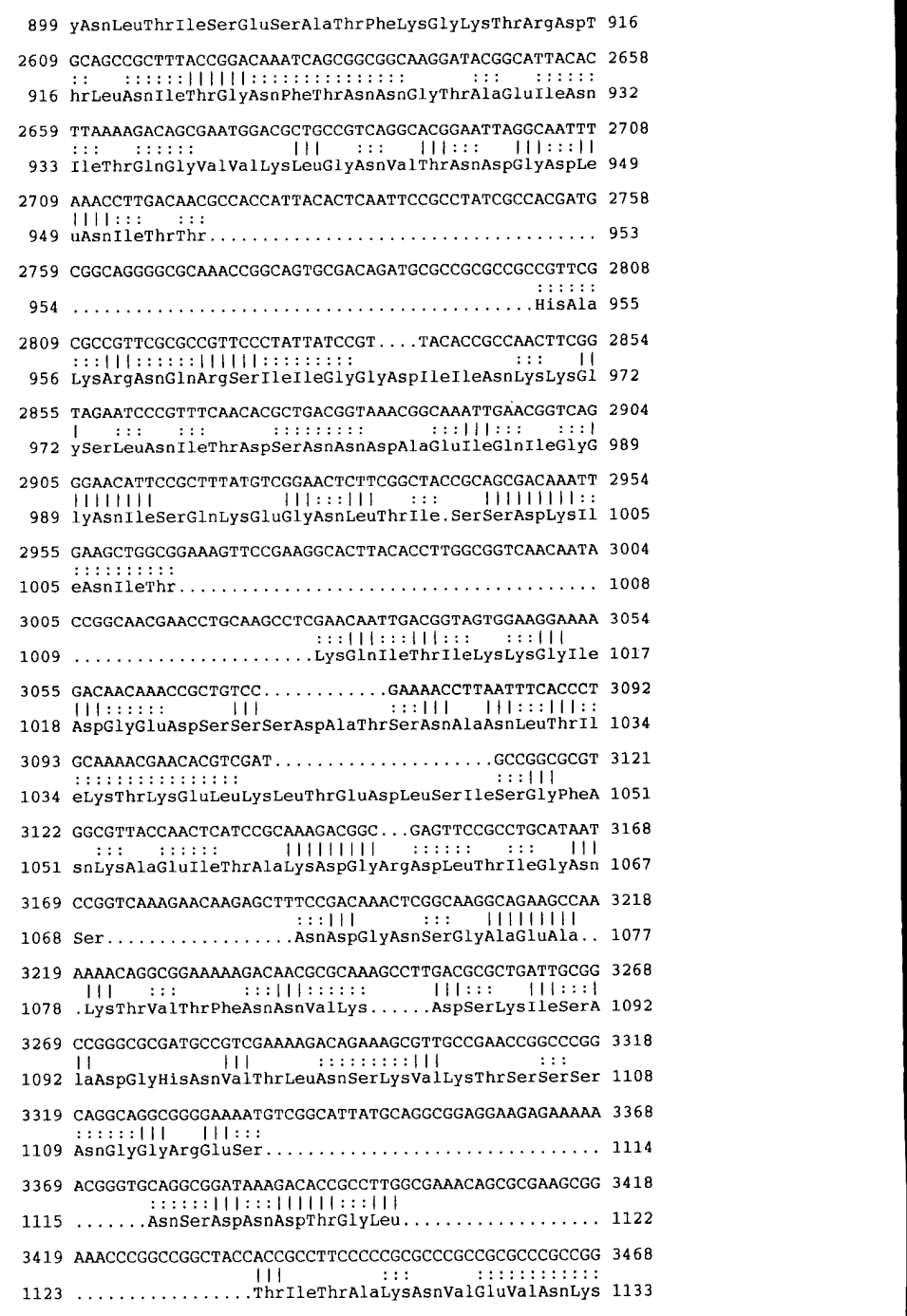
   1617 GCATTCGCTTTCGTTCCACCGTATTCAAAATACCGATGAAGGGGCGATGA 1666
        |                                 |||   :::|||      |
    614 y..............................ThrGlyLysGlyLeuAsnI 621

   1667 TTGTC............AACCACAATCAAGACAAAGAATCCACCGTTACC 1704
        ||:::           |||   :::::::::   ::::::|||::::::
    621 leIleSerSerValAsnAsnLeuThrHisAsnLeuSerGlyThrIleAsn 637

   1705 ATTACAGGCAATAAAGATATTGCTACAACCGGCAATAACAACAGCTTGGA 1754
        |||:::||||||      |||      |||
    638 IleSerGlyAsnIleThrIleAsnGlnThr..................... 647

   1755 TAGCAAAAAAGAAATTGCCTACAACGGTTGGTTTGGCGAGAAAGATACG. 1803
```


648 .ThrArgLysAsnThrSerTyr.....TrpGlnThrSerHisAspSerH 662
1804 ACCAAAACGAACGGGCGGCTCAACCTT 1830
662 isTrpAsnValSerAlaLeuAsnLeuGluThrGlyAlaAsnPheThrPhe 678
1831 GTTTACCAGCCCGCCGCAGAAGACCGCACCCTGCTGCTT..... 1869
679 IleLysTyrIleSerSerAsnSerLysGlyLeuThrThrGlnTyrArgSe 695
1870 .TCCGGCGGAACAAATTTAAACGGCAACATCACGCAAACAAACGGCAAAC 1918
695 rSerAlaGlyValAsnPheAsnGly.....ValAsnGlyAsnM 708
1919 TGTTTTTCAGCGGCAGACCAACACCGCACGCCTACAATCATTTAAACGAC 1968
708 etSerPhe..... 710
1969 CATTGGTCGCAAAAAGAGGGCATTCCTCGCGGGGAAATCGTGTGGGACAA 2018
711 AsnLeuLysGluGlyAlaLysValAsnPheLysLeuLysProAs 725
2019 CGACTGGATCAACCGCACA.....TTTAAAGCGGAAA 2050
725 nGluAsnMetAsnThrSerLysProLeuProIleArgPheLeuAla...A 741
2051 ACTTCCAAATTAAAGGCGGACAGGCGGTGGTT..... 2082
741 snIleThrAlaThrGlyGlyGlySerValPhePheAspIleTyrAlaAsn 757
2083 ...TCCCGCAATGTTGCCAAAGTGAAA...GGCGATTGGCATTTGAGCAA 2126
758 HisSerGlyArgGlyAlaGluLeuLysMetSerGluIleAsnIleSerAs 774
2127 TCACGCCCAAGCAGTTTTTGGTGTCGCACCGCATCAAAGCCACACAATCT 2176
774 nGlyAlaAsnPheThrLeu.....AsnSerHis..... 783
2177 GTACACGTTCGGACTGGACGGGTCTGACAAATTGTGTCGAAAAAACCATT 2226
784 ..ValArgGlyAspAspAlaPheLysIleAsn...LysAspLeuThrIle 798
2227 ACCGACGATAAAGTGATTGCTTCATTGACTAAGACCGACATCAGCGGCAA 2276
799 AsnAlaThrAsnSerAsnPheSerLeuArgGlnThrLysAspAspPheTy 815
2277 TGTCGATCTTGCCGATCACGCTCATTTAAATCTCACAGGGCTTGCCACAC 2326
815 rAspGlyTyrAlaArgAsnAlaIleAsnSerThrTyrAsnIleSerIleL 832
2327 TCAACGGCAATCTTAGTGCAAATGGCGATACACGTTATACAGTCAGCCAC 2376
832 euGlyGlyAsnValThrLeuGlyGly..... 840
2377 AACGCCACCCAAAACGGCAACCTTAGCCTCGTGGGCAATGCCCAAGCAAC 2426
841 GlnAsnSerSerSerSerIleThrGlyAsn.....IleTh 852
2427 ATTTAATCAAGCC.....ACATTAAACGGCAACACATCGGCTTCGG 2467
852 rIleGluLysAlaAlaAsnValThrLeuGluAlaAsnAsnAlaProAsnG 869
2468 GCAATGCTTCATTTAATCTAAGCGACCACGCCGTACAAAACGGCAGTCTG 2517
869 lnGln.....AsnIleArgAspArgValIleLysLeuGlySerLeu 882
2518 ACGCTTTCCGGCAACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACGG 2567
883 LeuValAsnGlySerLeuSerLeuThrGlyGluAsnAlaAspIleLysGl 899
2568 TAATGTCTCCCTAGCCGATAAGGCAGTATTCCAT.....TTTGAAA 2608


899 yAsnLeuThrIleSerGluSerAlaThrPheLysGlyLysThrArgAspT 916
2609 GCAGCCGCTTTACCGGACAAATCAGCGGCGGCAAGGATACGGCATTACAC 2658
916 hrLeuAsnIleThrGlyAsnPheThrAsnAsnGlyThrAlaGluIleAsn 932
2659 TTAAAAGACAGCGAATGGACGCTGCCGTCAGGCACGGAATTAGGCAATTT 2708
933 IleThrGlnGlyValValLysLeuGlyAsnValThrAsnAspGlyAspLe 949
2709 AAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCGCCACGATG 2758
949 uAsnIleThrThr..... 953
2759 CGGCAGGGGCGCAAACCGGCAGTGCGACAGATGCGCCGCGCCGCCGTTCG 2808
954 HisAla 955
2809 CGCCGTTCGCGCCGTTCCCTATTATCCGT....TACACCGCCAACTTCGG 2854
956 LysArgAsnGlnArgSerIleIleGlyGlyAspIleIleAsnLysLysGl 972
2855 TAGAATCCCGTTTCAACACGCTGACGGTAAACGGCAAATTGAACGGTCAG 2904
972 ySerLeuAsnIleThrAspSerAsnAsnAspAlaGluIleGlnIleGlyG 989
2905 GGAACATTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGACAAATT 2954
989 lyAsnIleSerGlnLysGluGlyAsnLeuThrIle.SerSerAspLysIl 1005
2955 GAAGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCGGTCAACAATA 3004
1005 eAsnIleThr..... 1008
3005 CCGGCAACGAACCTGCAAGCCTCGAACAATTGACGGTAGTGGAAGGAAAA 3054
1009 LysGlnIleThrIleLysLysGlyIle 1017
3055 GACAACAAACCGCTGTCC.....GAAAACCTTAATTTCACCCT 3092
1018 AspGlyGluAspSerSerSerAspAlaThrSerAsnAlaAsnLeuThrIl 1034
3093 GCAAAACGAACACGTCGAT.....GCCGGCGCGT 3121
1034 eLysThrLysGluLeuLysLeuThrGluAspLeuSerIleSerGlyPheA 1051
3122 GGCGTTACCAACTCATCCGCAAAGACGGC...GAGTTCCGCCTGCATAAT 3168
1051 snLysAlaGluIleThrAlaLysAspGlyArgAspLeuThrIleGlyAsn 1067
3169 CCGGTCAAAGAACAAGAGCTTTCCGACAAACTCGGCAAGGCAGAAGCCAA 3218
1068 Ser.....AsnAspGlyAsnSerGlyAlaGluAla.. 1077
3219 AAAACAGGCGGAAAAAGACAACGCGCAAAGCCTTGACGCGCTGATTGCGG 3268
1078 .LysThrValThrPheAsnAsnValLys.....AspSerLysIleSerA 1092
3269 CCGGGCGCGATGCCGTCGAAAAGACAGAAAGCGTTGCCGAACCGGCCCGG 3318
1092 laAspGlyHisAsnValThrLeuAsnSerLysValLysThrSerSerSer 1108
3319 CAGGCAGGCGGGGAAAATGTCGGCATTATGCAGGCGGAGGAAGAGAAAAA 3368
1109 AsnGlyGlyArgGluSer..... 1114
3369 ACGGGTGCAGGCGGATAAAGACACCGCCTTGGCGAAACAGCGCGAAGCGG 3418
1115 AsnSerAspAsnAspThrGlyLeu..... 1122
3419 AAACCCGGCCGGCTACCACCGCCTTCCCCCGCGCCCGCCGCGCCCGCCGG 3468
1123 ThrIleThrAlaLysAsnValGluValAsnLys 1133

```
     3469 GATTTGCCGCAACTGCAACCCCAACCGCAGCCCCAACCGCAGCGCGACCT 3518
          |||:::   :::|||
     1134 AspIleThrSerLeu..................................... 1138

     3519 GATCAGCCGTTATGCCAATAGCGGTTTGAGTGAATTTTCCGCCACGCTCA 3568
               :::   :::|||         ||||||      :::||| ::: 
     1139 .......LysThrValAsnIleThrAlaSerGluLysValThrThr.Thr 1152

     3569 ACAGCGTTTTCGCCGTACAGGACGAATTAGACCGCGTATTTGCCGAAGAC 3618
          ::::::  :::          ||||||      |||   ::: 
     1153 AlaGlySerThrIleAsnAlaThrAsnGlyLysAlaSerIleThrThrLy 1169

     3619 CGCCGCAACGCCGTTTGGACAAGCGGCATCCGGGACACCAAACACTACCG 3668
           ::::::  .      |||      ::::::::::|||   :::::::::: 
     1169 sThrGlyAspIleSerGlyThrIleSerGlyAsnThrValSerValSerA 1186

     3669 TTCGC...AAGATTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCAAA 3715
          ::         :::::::::   :::::::::|||      :::   ::: 
     1186 laThrValAspLeuThrThrLysSerGlySerLysIleGluAlaLysSer 1202

     3716 TCGGTATGCAGAAAAACCTCGGCAGCGGGCGCGTCGGCATCCTGTTTTCG 3765
          :::   ::::::   ||||||||:::|||:::   ::::::
     1203 GlyGluAlaAsnValThrSerAlaThrGlyThrIleGlyGly........ 1216

     3766 CACAACCGGACCGAAAACACCTTCGACGACGGCATCGGCA.......... 3805
           |||   :::   :::|||   ::::::||||||:::|||
     1217 .ThrIleSerGlyAsnThrValAsnValThrAlaAsnAlaGlyAspLeuT 1233

     3806 ........................ACTCGGCACGGCTTGCCCACGGCG 3829
                                   :::   :::|||      |||
     1233 hrValGlyAsnGlyAlaGluIleAsnAlaThrGluGlyAlaAlaThrLeu 1249

     3830 CCGTTTTCGGGCAATACGGCATCGACA......GGTTCTACATCGGCATC 3873
               :::||||||||||   :::|||      ||||||:::   :::|| 
     1250 ThrAlaThrGlyAsnThrLeuThrThrGluAlaGlySerSerIleThrSe 1266

     3874 AGCGCGGGCGCGGGTTTTAGCAGCGGCAGCCTTTCAGACGGCATCGGAGG 3923
          |:::::::::::::|||       |||      |||::::::|||   | 
     1266 rThrLysGlyGlnValAspLeuLeuAla......GlnAsnGlySerIleA 1281

     3924 CAAAATCCGCCGCCGCGTGCTGCATTACGGCATTCAGGCACGATACCGCG 3973
          ||   |||      |||:::   :::|||                  ||| 
     1281 laGlySerIleAsnAlaAlaAsnValThrLeuAsnThrThrGlyThrLeu 1297

     3974 CCGGTTTCGGCGGATTCGGCATCGAACCGCACATCGGCGCAACGCGCTAT 4023
             :::   |||   |||      :::   ||||||:::      :::|| 
     1298 ThrThrValAlaGlySerAspIleLysAlaThrSerGlyThrLeuValIl 1314

     4024 TTCGTCCAAA 4033
          |:::::::||
     1314 eAsnAlaLys 1317

seq_name: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-530-198-4

seq_documentation_block:
; Sequence 4, Application US/08530198
; Patent No. 5869065
;  GENERAL INFORMATION:
;    APPLICANT:  BARENKAMP, STEPHEN J
;    APPLICANT:  ST. GEME III, JOSEPH W
;    TITLE OF INVENTION:  HIGH MOLECULAR WEIGHT SURFACE PROTEINS
;    TITLE OF INVENTION:  OF NON-TYPEABLE HAEMOPHILUS
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Shoemaker and Mattare, Ltd
;      STREET:  2001 Jefferson Davis Hwy., 1203 Crystal Plaza
;      STREET:  Bldg. 1
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202-0286
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/530,198
;      FILING DATE:  13-DEC-1995
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BERKSTRESSER, JERRY W
;      REGISTRATION NUMBER:  22,651
;      REFERENCE/DOCKET NUMBER:  JWB-1186
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 415-0810
;      TELEFAX:  (703) 415-0813
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1477 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-530-198-4


alignment_scores:
            Quality:  282.50              Length:   1388
              Ratio:  0.428                 Gaps:     61
 Percent Similarity:  47.550   Percent Identity: 19.380

alignment_block:
 US-09-303-518D-649 x US-08-530-198-4   ..

 Align seg 1/1  to: US-08-530-198-4  from: 1  to: 1477

     175 GAAAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAA 224
         :::|||:::|||   :::::::|||::::::   :::|||:::::: 
     219 LysAsnGluGlyValIleSerValAsnGlyGlySerIleSerLeu..... 233

     225 CAAAAAAGGGGAGTTGGTCGGCAAATCAATGACAAAAGCC.......... 264
                     |||:::|||:::::::::|||   :::
     234 .............LeuAlaGlyGlnLysIleThrIleSerAspIleIleA 246

     265 ..CCGATGATTGATTTTTCTGTGGTGTCGCGTAACGGCGTGGCGGCATTG 312
           |||   |||   :::|||:::::::::   ::::::   |||::: 
     246 snProThrIleThrTyrSerIleAlaAlaProGluAsnGluAlaValAsn 262

     313 GTGGGCGATCAATATATTGTGAGCGTGGCACATAACGGCGGCTATAACAA 362
         :::||||||                :::      :::||||||      || 
     263 LeuGlyAsp..............IlePheAlaLysGlyGlyAsnIleAs 274

     363 CGTT..........................GATTTTGGTGCGGAAGGAA 385
         ||||                         :::::::|||:::::: 
     274 nValArgAlaAlaThrIleArgAsnGlnGlyLysLeuSerAlaAspServ 291

     386 GAAATCCCGATCAACATCGTTTTACTTATAAAATTGTGAAACGGAATAAT 435
           :::   |||:::                        :::|||:::::::::: 
     291 alSerLysAspLysSerGlyAsnIleValLeuSerAlaLysGluGlyGlu 307

     436 TATAAAGCAGGG..............ACTAAAGGCCATCCTTATGGCGG 470
           :::   |||                :::   :::      ||||| 
     308 AlaGluIleGlyGlyValIleSerAlaGlnAsnGlnGlnAlaLysGlyGl 324

     471 CGATTATCATATGCCGCGTTTGCATAAATTTGTCACAGATGCAGAACCTG 520
         |      :::          :::      :::      :
     324 yLysLeuMetIleThrGlyAspLysValThrLeuLysThrGlyAlaValI 341

     521 TTGAAATGACCAGTTATATGGATGGGCGGAAATATATCGATCAAAATAAT 570
         ::::::::::::::         |||:::   |||:::      ::::::
     341 leAspLeuSerGlyLysGluGlyGlyGluThrTyrLeuGlyGlyAspGlu 357

     571 TACCCTGACCGTGTTCGTATTGGGGCAGGCAGGCAATATTGGCGATCTGA 620
                               |||   |||::::::      :::
```





```
    1298 ThrThrValAlaGlySerAspIleLysAlaThrSerGlyThrLeuValIl 1314

    4024 TTCGTCCAAA 4033
         |::::::|||
    1314 eAsnAlaLys 1317

seq_name: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-469-880-4

seq_documentation_block:
; Sequence 4, Application US/08469880
; Patent No. 5876733
;  GENERAL INFORMATION:
;    APPLICANT:  Barenkamp, Stephen J.
;    TITLE OF INVENTION:  High Molecular Weight Surface Proteins
;    TITLE OF INVENTION:  of No. 5876733-Typeable Haemophilus
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Shoemaker and Mattare, Ltd.
;      STREET:  2001 Jefferson Davis Hwy., 1203 Crystal Plaza
;      STREET:  Bldg. 1
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202-0286
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/469,880
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9205704.1
;      FILING DATE:  16-MAR-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US PCT/US93/02166
;      FILING DATE:  16-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/302,832
;      FILING DATE:  16-SEP-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Berkstresser, Jerry W
;      REGISTRATION NUMBER:  22,651
;      REFERENCE/DOCKET NUMBER:  1038-516 MIS:vg
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 415-0810
;      TELEFAX:  (703) 415-0813
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1477 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-469-880-4


alignment_scores:
            Quality:  282.50              Length:   1388
              Ratio:  0.428                Gaps:     61
 Percent Similarity:  47.550   Percent Identity: 19.380

alignment_block:
 US-09-303-518D-649 x US-08-469-880-4   ..

 Align seg 1/1  to: US-08-469-880-4  from: 1  to: 1477

     175 GAAAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAA 224
         :::|||:::|||   :::::::|||::::::   :::|||::::::
     219 LysAsnGluGlyValIleSerValAsnGlyGlySerIleSerLeu..... 233

     225 CAAAAAAGGGGAGTTGGTCGGCAAATCAATGACAAAAGCC.......... 264

                 |||:::|||:::::::::|||   :::
 234 .............LeuAlaGlyGlnLysIleThrIleSerAspIleIleA 246

 265 ..CCGATGATTGATTTTTCTGTGGTGTCGCGTAACGGCGTGGCGGCATTG 312
       |||   |||   :::|||:::::::::   ::::::   |||:::
 246 snProThrIleThrTyrSerIleAlaAlaProGluAsnGluAlaValAsn 262

 313 GTGGGCGATCAATATATTGTGAGCGTGGCACATAACGGCGGCTATAACAA 362
     :::||||||               :::      :::||||||      ||
 263 LeuGlyAsp..............IlePheAlaLysGlyGlyAsnIleAs 274

 363 CGTT........................GATTTTGGTGCGGAAGGAA 385
     ||||                       ::::::|||::::::
 274 nValArgAlaAlaThrIleArgAsnGlnGlyLysLeuSerAlaAspSerV 291

 386 GAAATCCCGATCAACATCGTTTTACTTATAAAATTGTGAAACGGAATAAT 435
       :::   |||:::                    :::|||::::::::::
 291 alSerLysAspLysSerGlyAsnIleValLeuSerAlaLysGluGlyGlu 307

 436 TATAAAGCAGGG..............ACTAAAGGCCATCCTTATGGCGG 470
       :::   |||                 :::   :::      |||||
 308 AlaGluIleGlyGlyValIleSerAlaGlnAsnGlnGlnAlaLysGlyGl 324

 471 CGATTATCATATGCCGCGTTTGCATAAATTTGTCACAGATGCAGAACCTG 520
     |         :::                 :::      :::      :
 324 yLysLeuMetIleThrGlyAspLysValThrLeuLysThrGlyAlaValI 341

 521 TTGAAATGACCAGTTATATGGATGGGCGGAAATATATCGATCAAAATAAT 570
     ::::::::::::::        |||:::   |||:::      ::::::
 341 leAspLeuSerGlyLysGluGlyGlyGluThrTyrLeuGlyGlyAspGlu 357

 571 TACCCTGACCGTGTTCGTATTGGGGCAGGCAGGCAATATTGGCGATCTGA 620
                         |||   |||::::::      :::
 358 Arg.................GlyGluGlyLysAsnGlyIleGlnLeuAl 368

 621 TGAAGATGAGCCCAATAACCGCGAAAGTTCATATCATATTGCAAGT.... 666
       :::           ::::::   |||:::   ::::::::::::
 368 aLysLysThrSerLeuGluLysGlySerThrIleAsnValSerGlyLysG 385

 667 .............GCGTATTCTTGG............CTCGTTGGTGGC 690
                 |||      |||            |||:::   |||
 385 luLysGlyGlyPheAlaIleValTrpGlyAspIleAlaLeuIleAspGly 401

 691 AATACCTTTGCACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAG 740
     |||      ||||||   |||||||||
 402 AsnIleAsnAlaGln...GlySerGly..................... 409

 741 TGAAAAAATTAAACATAGCCCATATGGTTTTTTACCAACAGGAGGCTCAT 790
                                 :::      |||||||||
 410 .......................AspIleAlaLysThrGlyGlyPheV 418

 791 TTGGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAG 840
             ||||||      :::||||||   |||
 418 alGluThrSerGlyHisAspLeuPheIleLysAsp.............. 429

 841 TGGTTAATTAATGGGGTATTGCAAACGGGCAACCCCTATATAGGAAAAAG 890
             |||:::::::::::::::
 430 .........AsnAlaIleValAspAla..................... 435

 891 CAATGGCTTCCAGCTGGTTCGTAAAGATTGGTTCTATGATGAAATCTTTG 940
                              |||:::|||:::   |||
 436 .....................LysGluTrpLeuLeuAsp......... 441

 941 CTGGAGATACCCATTCAGTATTCTACGAACCACGTCAAAATGGGAAATAC 990
         |||:::   |||:::      |||             ::::::
 442 ..PheAspAsnValSerIleAsnAlaGluAspProLeuPheAsnAsnThr 457

 991 TCTTTTAACGACGAT...AATAATGGCACAGGAAAAATCAATGCCAAACA 1037
     :::::::||||||:::      :::||||||||||:::   :::
 458 GlyIleAsnAspGluPheProThrGlyThrGlyGluAlaSerAspProLy 474

1038 TGAACACAATTCTCTGCCTAATAGATTAAAAACACGAACCGTTCAATTGT 1087
      :::::::::::|||   :::   |||   :::   |||::::::   :
```



```
2759 CGGCAGGGGCGCAAACCGGCAGTGCGACAGATGCGCCGCGCCGCCGTTCG 2808
                                                  ::::::
 954 ..........................................HisAla 955

2809 CGCCGTTCGCGCCGTTCCCTATTATCCGT....TACACCGCCAACTTCGG 2854
     :::|||::::::||||||:::::::::          :::   ||
 956 LysArgAsnGlnArgSerIleIleGlyGlyAspIleIleAsnLysLysGl 972

2855 TAGAATCCCGTTTCAACACGCTGACGGTAAACGGCAAATTGAACGGTCAG 2904
     |   :::   :::      :::::::::      :::|||:::   :::|
 972 ySerLeuAsnIleThrAspSerAsnAsnAspAlaGluIleGlnIleGlyG 989

2905 GGAACATTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGACAAATT 2954
     ||||||||          |||:::|||   :::    ||||||||::
 989 lyAsnIleSerGlnLysGluGlyAsnLeuThrIle.SerSerAspLysIl 1005

2955 GAAGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCGGTCAACAATA 3004
     ::::::::::
1005 eAsnIleThr..................................... 1008

3005 CCGGCAACGAACCTGCAAGCCTCGAACAATTGACGGTAGTGGAAGGAAAA 3054
                          :::|||:::|||:::   :::|||
1009 .......................LysGlnIleThrIleLysLysGlyIle 1017

3055 GACAACAAACCGCTGTCC............GAAAACCTTAATTTCACCCT 3092
     |||::::::     |||          :::|||   |||:::||||::
1018 AspGlyGluAspSerSerSerAspAlaThrSerAsnAlaAsnLeuThrIl 1034

3093 GCAAAACGAACACGTCGAT.....................GCCGGCGCGT 3121
     ::::::::::::::::                        :::|||
1034 eLysThrLysGluLeuLysLeuThrGluAspLeuSerIleSerGlyPheA 1051

3122 GGCGTTACCAACTCATCCGCAAAGACGGC...GAGTTCCGCCTGCATAAT 3168
       :::   ::::::     |||||||||   ::::::   :::   |||
1051 snLysAlaGluIleThrAlaLysAspGlyArgAspLeuThrIleGlyAsn 1067

3169 CCGGTCAAAGAACAAGAGCTTTCCGACAAACTCGGCAAGGCAGAAGCCAA 3218
                          :::|||      :::   |||||||||
1068 Ser.................AsnAspGlyAsnSerGlyAlaGluAla.. 1077

3219 AAAACAGGCGGAAAAAGACAACGCGCAAAGCCTTGACGCGCTGATTGCGG 3268
      |||   :::      :::|||::::::      |||:::   |||:::|
1078 .LysThrValThrPheAsnAsnValLys......AspSerLysIleSerA 1092

3269 CCGGGCGCGATGCCGTCGAAAAGACAGAAAGCGTTGCCGAACCGGCCCGG 3318
     ||            |||      :::::::::|||          :::
1092 laAspGlyHisAsnValThrLeuAsnSerLysValLysThrSerSerSer 1108

3319 CAGGCAGGCGGGGAAAATGTCGGCATTATGCAGGCGGAGGAAGAGAAAAA 3368
     :::::::|||   |||:::
1109 AsnGlyGlyArgGluSer.............................. 1114

3369 ACGGGTGCAGGCGGATAAAGACACCGCCTTGGCGAAACAGCGCGAAGCGG 3418
            ::::::|||:::||||||:::|||
1115 .......AsnSerAspAsnAspThrGlyLeu................... 1122

3419 AAACCCGGCCGGCTACCACCGCCTTCCCCCGCGCCCGCCGCGCCCGCCGG 3468
                      |||          :::      ::::::::::::
1123 .................ThrIleThrAlaLysAsnValGluValAsnLys 1133

3469 GATTTGCCGCAACTGCAACCCCAACCGCAGCCCCAACCGCAGCGCGACCT 3518
     |||:::   :::|||
1134 AspIleThrSerLeu................................... 1138

3519 GATCAGCCGTTATGCCAATAGCGGTTTGAGTGAATTTTCCGCCACGCTCA 3568
            :::   :::|||          ||||||      :::||| :::
1139 .......LysThrValAsnIleThrAlaSerGluLysValThrThr.Thr 1152

3569 ACAGCGTTTTCGCCGTACAGGACGAATTAGACCGCGTATTTGCCGAAGAC 3618
     ::::::   :::          ||||||      |||   :::
1153 AlaGlySerThrIleAsnAlaThrAsnGlyLysAlaSerIleThrThrLy 1169

3619 CGCCGCAACGCCGTTTGGACAAGCGGCATCCGGGACACCAAACACTACCG 3668
```

```
          ::::::          |||      :::::::::|||   ::::::::::
1169 sThrGlyAspIleSerGlyThrIleSerGlyAsnThrValSerValSerA 1186

3669 TTCGC...AAGATTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCAAA 3715
     ::         :::::::::   :::::::::|||      :::   :::
1186 laThrValAspLeuThrThrLysSerGlySerLysIleGluAlaLysSer 1202

3716 TCGGTATGCAGAAAAACCTCGGCAGCGGGCGCGTCGGCATCCTGTTTTCG 3765
     :::   ::::::   |||||||||:::|||:::   ::::::
1203 GlyGluAlaAsnValThrSerAlaThrGlyThrIleGlyGly........ 1216

3766 CACAACCGGACCGAAAACACCTTCGACGACGGCATCGGCA.......... 3805
      |||   :::   :::|||   ::::::||||||:::|||
1217 .ThrIleSerGlyAsnThrValAsnValThrAlaAsnAlaGlyAspLeuT 1233

3806 ........................ACTCGGCACGGCTTGCCCACGGCG 3829
                               :::   :::|||      |||
1233 hrValGlyAsnGlyAlaGluIleAsnAlaThrGluGlyAlaAlaThrLeu 1249

3830 CCGTTTTCGGGCAATACGGCATCGACA......GGTTCTACATCGGCATC 3873
           :::|||||||||   :::|||      ||||||:::   :::||
1250 ThrAlaThrGlyAsnThrLeuThrThrGluAlaGlySerSerIleThrSe 1266

3874 AGCGCGGGCGCGGGTTTTAGCAGCGGCAGCCTTTCAGACGGCATCGGAGG 3923
     |:::::::::::::|||          |||      |||:::::::|||   |
1266 rThrLysGlyGlnValAspLeuLeuAla......GlnAsnGlySerIleA 1281

3924 CAAAATCCGCCGCCGCGTGCTGCATTACGGCATTCAGGCACGATACCGCG 3973
     ||   |||      |||:::   :::|||                |||
1281 laGlySerIleAsnAlaAlaAsnValThrLeuAsnThrThrGlyThrLeu 1297

3974 CCGGTTTCGGCGGATTCGGCATCGAACCGCACATCGGCGCAACGCGCTAT 4023
        :::   |||   |||      :::   ||||||:::      :::||
1298 ThrThrValAlaGlySerAspIleLysAlaThrSerGlyThrLeuValIl 1314

4024 TTCGTCCAAA 4033
     |:::::::||
1314 eAsnAlaLys 1317
```

seq_name: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-728-470-4

```
seq_documentation_block:
; Sequence 4, Application US/08728470
; Patent No. 5928651
;  GENERAL INFORMATION:
;    APPLICANT:  Barenkamp, Stephen J
;    TITLE OF INVENTION:  High Molecular Weight Surface Proteins
;    TITLE OF INVENTION:  of No. 5928651-Typeable Haemophilus
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Shoemaker and Mattare, Ltd.
;      STREET:  2001 Jefferson Davis Hwy., 1203 Crystal Plaza
;      STREET:  Bldg. 1
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202-0286
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/728,470
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/302,832
;      FILING DATE:  16-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US PCT/US93/02166
;      FILING DATE:  16-MAR-1993
;    PRIOR APPLICATION DATA:
```

```
;      APPLICATION NUMBER:  GB 9205704.1
;      FILING DATE:  16-MAR-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Berkstresser, Jerry W
;      REGISTRATION NUMBER:  22,651
;      REFERENCE/DOCKET NUMBER:  1038-633
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 415-0810
;      TELEFAX:  (703) 415-0813
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1477 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-728-470-4


alignment_scores:
           Quality:  282.50              Length:   1388
             Ratio:  0.428                Gaps:     61
 Percent Similarity:  47.550   Percent Identity: 19.380

alignment_block:
 US-09-303-518D-649 x US-08-728-470-4   ..

 Align seg 1/1  to: US-08-728-470-4  from: 1  to: 1477

     175 GAAAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAA 224
         :::|||:::|||   :::::::|||:::::::   :::|||::::::
     219 LysAsnGluGlyValIleSerValAsnGlyGlySerIleSerLeu..... 233

     225 CAAAAAAGGGGAGTTGGTCGGCAAATCAATGACAAAAGCC.......... 264
                    |||:::|||::::::::::|||   :::
     234 .............LeuAlaGlyGlnLysIleThrIleSerAspIleIleA 246

     265 ..CCGATGATTGATTTTTCTGTGGTGTCGCGTAACGGCGTGGCGGCATTG 312
           |||   |||   :::|||::::::::::   ::::::   |||:::
     246 snProThrIleThrTyrSerIleAlaAlaProGluAsnGluAlaValAsn 262

     313 GTGGGCGATCAATATATTGTGAGCGTGGCACATAACGGCGGCTATAACAA 362
         :::||||||               :::      :::|||||||     ||
     263 LeuGlyAsp..............IlePheAlaLysGlyGlyAsnIleAs 274

     363 CGTT.........................GATTTTGGTGCGGAAGGAA 385
         ||||                        :::::::|||::::::
     274 nValArgAlaAlaThrIleArgAsnGlnGlyLysLeuSerAlaAspServ 291

     386 GAAATCCCGATCAACATCGTTTTACTTATAAAATTGTGAAACGGAATAAT 435
           :::   |||:::                     :::|||::::::::::
     291 alSerLysAspLysSerGlyAsnIleValLeuSerAlaLysGluGlyGlu 307

     436 TATAAAGCAGGG..............ACTAAAGGCCATCCTTATGGCGG 470
           :::   |||                 :::   :::      |||||
     308 AlaGluIleGlyGlyValIleSerAlaGlnAsnGlnGlnAlaLysGlyGl 324

     471 CGATTATCATATGCCGCGTTTGCATAAATTTGTCACAGATGCAGAACCTG 520
         |       :::                 :::      :::      :
     324 yLysLeuMetIleThrGlyAspLysValThrLeuLysThrGlyAlaValI 341

     521 TTGAAATGACCAGTTATATGGATGGGCGGAAATATATCGATCAAAATAAT 570
         ::::::::::::::          |||:::   |||:::      ::::::
     341 leAspLeuSerGlyLysGluGlyGlyGluThrTyrLeuGlyGlyAspGlu 357

     571 TACCCTGACCGTGTTCGTATTGGGGCAGGCAGGCAATATTGGCGATCTGA 620
                       |||   |||::::::      :::
     358 Arg.................GlyGluGlyLysAsnGlyIleGlnLeuAl 368

     621 TGAAGATGAGCCCAATAACCGCGAAAGTTCATATCATATTGCAAGT.... 666
           :::             ::::::   |||:::   ::::::::::::
     368 aLysLysThrSerLeuGluLysGlySerThrIleAsnValSerGlyLysG 385

     667 .............GCGTATTCTTGG............CTCGTTGGTGGC 690

                       |||      |||           |||:::   |||
     385 luLysGlyGlyPheAlaIleValTrpGlyAspIleAlaLeuIleAspGly 401

     691 AATACCTTTGCACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAG 740
         |||       ||||||   ||||||||
     402 AsnIleAsnAlaGln...GlySerGly..................... 409

     741 TGAAAAAATTAAACATAGCCCATATGGTTTTTTACCAACAGGAGGCTCAT 790
                                        :::       |||||||||
     410 .......................AspIleAlaLysThrGlyGlyPheV 418

     791 TTGGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAG 840
               ||||||      :::||||||   |||
     418 alGluThrSerGlyHisAspLeuPheIleLysAsp.............. 429

     841 TGGTTAATTAATGGGGTATTGCAAACGGGCAACCCCTATATAGGAAAAAG 890
              |||:::::::::::::::
     430 .........AsnAlaIleValAspAla..................... 435

     891 CAATGGCTTCCAGCTGGTTCGTAAAGATTGGTTCTATGATGAAATCTTTG 940
                                |||:::|||:::   |||
     436 .....................LysGluTrpLeuLeuAsp.......... 441

     941 CTGGAGATACCCATTCAGTATTCTACGAACCACGTCAAAATGGGAAATAC 990
           |||:::   |||:::      |||           ::::::
     442 ..PheAspAsnValSerIleAsnAlaGluAspProLeuPheAsnAsnThr 457

     991 TCTTTTAACGACGAT...AATAATGGCACAGGAAAAATCAATGCCAAACA 1037
         ::::::|||||:::      :::|||||||||:::   :::
     458 GlyIleAsnAspGluPheProThrGlyThrGlyGluAlaSerAspProLy 474

    1038 TGAACACAATTCTCTGCCTAATAGATTAAAAACACGAACCGTTCAATTGT 1087
          :::::::::::::|||   :::   |||   :::   |||::::::   :
     474 sLysAsnSerGluLeuLysThrThrLeuThrAsnThrThrIleSerAsnT 491

    1088 TT.....................AATGTTTCTTTATCCGAGACAGCAAGA 1116
         ::                        |||::::::   |||:::
     491 yrLeuLysAsnAlaTrpThrMetAsnIleThrAlaSerArgLysLeuThr 507

    1117 GAACCTGTTTATCATGCTGCAGGTGGTGTCAACAGTTATCGACCCAGACT 1166
                                        |||||||||           ::
     508 .........................ValAsnSer.........SerIl 512

    1167 GAATAATGGAGAAAATATTTCCTTTATTGACGAAGGAAAAGGCGAATTGA 1216
         :|||   |||:::|||      :::|||   ::::::||||||
     512 eAsnIleGlySerAsnSerHisLeuIleLeuHisSerLysGly....... 526

    1217 TACTTACCAGCAACATCAATCAAGGTGCTGGAGGATTATATTTCCAAGGA 1266
                              |||   :::||||||:::   ::::::|||
     527 ...................GlnArgGlyGlyGlyValGlnIleAspGly 536

    1267 GATTTTACGGTCTCGCCTGAAAATAACGAAACTTGGCAAGGCGCGGGCGT 1316
         |||:::|||   :::       |||          :::::::|||:::   ||
     537 AspIleThrSerLysGlyGlyAsnLeuThrIleTyrSerGlyGlyTrpVa 553

    1317 TCATATCAGTGAAGACAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAA 1366
         |   :::   ::::::   |||:::
     553 lAspValHisLysAsnIleThrLeu........................ 561

    1367 ACGACCGCCTGTCCAAAATCGGCAAAGGCACGCTGCACGTTCAAGCCAAA 1416

     561 ................................................. 561

    1417 GGGGAAAACCAAGGCTCGATCAGCGTGGGCGACGGTACAGTCATTTTGGA 1466
               :::||||||   :::::::::      ::::::|||   :::
     562 ......AspGlnGlyPheLeuAsnIleThrAlaAlaSerValAlaPhe.. 575

    1467 TCAGCAGGCAGACGATAAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCT 1516
          :::   ::::::::::|||::::::      |||   ::::::|||
     576 .GluGlyGlyAsnAsnLysAlaArgAspAlaAlaAsnAlaLysIle.... 590

    1517 TGGTCAGCGGCAGGGGTACGGTGCAACTGAATGCCGATAATCAGTTCAAC 1566
               ::::::|||||||||   ::::::::::::::::::
```


591 ValAlaGlnGlyThrValThrIleThrGlyGluGlyLys..... 603
1567 CCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAACGG 1616
604 AspPheArgAlaAsnAsnValSerLeuAsnGl 614
1617 GCATTCGCTTTCGTTCCACCGTATTCAAAATACCGATGAAGGGGCGATGA 1666
614 y.....ThrGlyLysGlyLeuAsnI 621
1667 TTGTC.....AACCACAATCAAGACAAAGAATCCACCGTTACC 1704
621 leIleSerSerValAsnAsnLeuThrHisAsnLeuSerGlyThrIleAsn 637
1705 ATTACAGGCAATAAAGATATTGCTACAACCGGCAATAACAACAGCTTGGA 1754
638 IleSerGlyAsnIleThrIleAsnGlnThr..... 647
1755 TAGCAAAAAAGAAATTGCCTACAACGGTTGGTTTGGCGAGAAAGATACG. 1803
648 .ThrArgLysAsnThrSerTyr.....TrpGlnThrSerHisAspSerH 662
1804 ACCAAAACGAACGGGCGGCTCAACCTT 1830
662 isTrpAsnValSerAlaLeuAsnLeuGluThrGlyAlaAsnPheThrPhe 678
1831 GTTTACCAGCCCGCCGCAGAAGACCGCACCCTGCTGCTT..... 1869
679 IleLysTyrIleSerSerAsnSerLysGlyLeuThrThrGlnTyrArgSe 695
1870 .TCCGGCGGAACAAATTTAAACGGCAACATCACGCAAACAAACGGCAAAC 1918
695 rSerAlaGlyValAsnPheAsnGly.....ValAsnGlyAsnM 708
1919 TGTTTTTCAGCGGCAGACCAACACCGCACGCCTACAATCATTTAAACGAC 1968
708 etSerPhe..... 710
1969 CATTGGTCGCAAAAAGAGGGCATTCCTCGCGGGGAAATCGTGTGGGACAA 2018
711 AsnLeuLysGluGlyAlaLysValAsnPheLysLeuLysProAs 725
2019 CGACTGGATCAACCGCACA.....TTTAAAGCGGAAA 2050
725 nGluAsnMetAsnThrSerLysProLeuProIleArgPheLeuAla...A 741
2051 ACTTCCAAATTAAAGGCGGACAGGCGGTGGTT..... 2082
741 snIleThrAlaThrGlyGlyGlySerValPhePheAspIleTyrAlaAsn 757
2083 ...TCCCGCAATGTTGCCAAAGTGAAA...GGCGATTGGCATTTGAGCAA 2126
758 HisSerGlyArgGlyAlaGluLeuLysMetSerGluIleAsnIleSerAs 774
2127 TCACGCCCAAGCAGTTTTTGGTGTCGCACCGCATCAAAGCCACACAATCT 2176
774 nGlyAlaAsnPheThrLeu.....AsnSerHis..... 783
2177 GTACACGTTCGGACTGGACGGGTCTGACAAATTGTGTCGAAAAAACCATT 2226
784 ..ValArgGlyAspAspAlaPheLysIleAsn...LysAspLeuThrIle 798
2227 ACCGACGATAAAGTGATTGCTTCATTGACTAAGACCGACATCAGCGGCAA 2276
799 AsnAlaThrAsnSerAsnPheSerLeuArgGlnThrLysAspAspPheTy 815
2277 TGTCGATCTTGCCGATCACGCTCATTTAAATCTCACAGGGCTTGCCACAC 2326
815 rAspGlyTyrAlaArgAsnAlaIleAsnSerThrTyrAsnIleSerIleL 832
2327 TCAACGGCAATCTTAGTGCAAATGGCGATACACGTTATACAGTCAGCCAC 2376
832 euGlyGlyAsnValThrLeuGlyGly..... 840
2377 AACGCCACCCAAAACGGCAACCTTAGCCTCGTGGGCAATGCCCAAGCAAC 2426
841 GlnAsnSerSerSerSerIleThrGlyAsn.....IleTh 852
2427 ATTTAATCAAGCC.....ACATTAAACGGCAACACATCGGCTTCGG 2467
852 rIleGluLysAlaAlaAsnValThrLeuGluAlaAsnAsnAlaProAsnG 869
2468 GCAATGCTTCATTTAATCTAAGCGACCACGCCGTACAAAACGGCAGTCTG 2517
869 lnGln.....AsnIleArgAspArgValIleLysLeuGlySerLeu 882
2518 ACGCTTTCCGGCAACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACGG 2567
883 LeuValAsnGlySerLeuSerLeuThrGlyGluAsnAlaAspIleLysGl 899
2568 TAATGTCTCCCTAGCCGATAAGGCAGTATTCCAT.....TTTGAAA 2608
899 yAsnLeuThrIleSerGluSerAlaThrPheLysGlyLysThrArgAspT 916
2609 GCAGCCGCTTTACCGGACAAATCAGCGGCGGCAAGGATACGGCATTACAC 2658
916 hrLeuAsnIleThrGlyAsnPheThrAsnAsnGlyThrAlaGluIleAsn 932
2659 TTAAAAGACAGCGAATGGACGCTGCCGTCAGGCACGGAATTAGGCAATTT 2708
933 IleThrGlnGlyValValLysLeuGlyAsnValThrAsnAspGlyAspLe 949
2709 AAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCGCCACGATG 2758
949 uAsnIleThrThr..... 953
2759 CGGCAGGGGCGCAAACCGGCAGTGCGACAGATGCGCCGCGCCGCCGTTCG 2808
954 HisAla 955
2809 CGCCGTTCGCGCCGTTCCCTATTATCCGT....TACACCGCCAACTTCGG 2854
956 LysArgAsnGlnArgSerIleIleGlyGlyAspIleIleAsnLysLysGl 972
2855 TAGAATCCCGTTTCAACACGCTGACGGTAAACGGCAAATTGAACGGTCAG 2904
972 ySerLeuAsnIleThrAspSerAsnAsnAspAlaGluIleGlnIleGlyG 989
2905 GGAACATTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGACAAATT 2954
989 lyAsnIleSerGlnLysGluGlyAsnLeuThrIle.SerSerAspLysIl 1005
2955 GAAGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCGGTCAACAATA 3004
1005 eAsnIleThr..... 1008
3005 CCGGCAACGAACCTGCAAGCCTCGAACAATTGACGGTAGTGGAAGGAAAA 3054
1009 LysGlnIleThrIleLysLysGlyIle 1017
3055 GACAACAAACCGCTGTCC.....GAAAACCTTAATTTCACCCT 3092
1018 AspGlyGluAspSerSerSerAspAlaThrSerAsnAlaAsnLeuThrIl 1034
3093 GCAAAACGAACACGTCGAT.....GCCGGCGCGT 3121
1034 eLysThrLysGluLeuLysLeuThrGluAspLeuSerIleSerGlyPheA 1051
3122 GGCGTTACCAACTCATCCGCAAAGACGGC...GAGTTCCGCCTGCATAAT 3168
1051 snLysAlaGluIleThrAlaLysAspGlyArgAspLeuThrIleGlyAsn 1067
3169 CCGGTCAAAGAACAAGAGCTTTCCGACAAACTCGGCAAGGCAGAAGCCAA 3218
1068 Ser.....AsnAspGlyAsnSerGlyAlaGluAla.. 1077

```
 3219 AAAACAGGCGGAAAAAGACAACGCGCAAAGCCTTGACGCGCTGATTGCGG 3268
      |||   :::      :::|||::::::      |||:::  |||:::|
 1078 .LysThrValThrPheAsnAsnValLys......AspSerLysIleSerA 1092

 3269 CCGGGCGCGATGCCGTCGAAAAGACAGAAAGCGTTGCCGAACCGGCCCGG 3318
      ||           |||      ::::::::::|||         :::.
 1092 laAspGlyHisAsnValThrLeuAsnSerLysValLysThrSerSerSer 1108

 3319 CAGGCAGGCGGGGAAAATGTCGGCATTATGCAGGCGGAGGAAGAGAAAAA 3368
      ::::::|||  |||:::
 1109 AsnGlyGlyArgGluSer..................................... 1114

 3369 ACGGGTGCAGGCGGATAAAGACACCGCCTTGGCGAAACAGCGCGAAGCGG 3418
             ::::::|||:::||||||:::|||
 1115 .......AsnSerAspAsnAspThrGlyLeu..................... 1122

 3419 AAACCCGGCCGGCTACCACCGCCTTCCCCCGCGCCCGCCGCGCCCGCCGG 3468
                      |||         :::      ::::::::::::
 1123 .................ThrIleThrAlaLysAsnValGluValAsnLys 1133

 3469 GATTTGCCGCAACTGCAACCCCAACCGCAGCCCCAACCGCAGCGCGACCT 3518
      |||:::   :::|||
 1134 AspIleThrSerLeu..................................... 1138

 3519 GATCAGCCGTTATGCCAATAGCGGTTTGAGTGAATTTTCCGCCACGCTCA 3568
             :::   :::|||       ||||||      :::||| :::
 1139 .......LysThrValAsnIleThrAlaSerGluLysValThrThr.Thr 1152

 3569 ACAGCGTTTTCGCCGTACAGGACGAATTAGACCGCGTATTTGCCGAAGAC 3618
      ::::::   :::         ||||||      |||   :::
 1153 AlaGlySerThrIleAsnAlaThrAsnGlyLysAlaSerIleThrThrLy 1169

 3619 CGCCGCAACGCCGTTTGGACAAGCGGCATCCGGGACACCAAACACTACCG 3668
      ::::::         |||      :::::::::|||   ::::::::::
 1169 sThrGlyAspIleSerGlyThrIleSerGlyAsnThrValSerValSerA 1186

 3669 TTCGC...AAGATTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCAAA 3715
      ::         :::::::::   ::::::::::|||      :::  :::
 1186 laThrValAspLeuThrThrLysSerGlySerLysIleGluAlaLysSer 1202

 3716 TCGGTATGCAGAAAAACCTCGGCAGCGGGCGCGTCGGCATCCTGTTTTCG 3765
      :::   ::::::   |||||||||:::|||:::   ::::::
 1203 GlyGluAlaAsnValThrSerAlaThrGlyThrIleGlyGly........ 1216

 3766 CACAACCGGACCGAAAACACCTTCGACGACGGCATCGGCA.......... 3805
       |||   :::   :::|||   ::::::||||||:::|||
 1217 .ThrIleSerGlyAsnThrValAsnValThrAlaAsnAlaGlyAspLeuT 1233

 3806 .........................ACTCGGCACGGCTTGCCCACGGCG 3829
                               :::   :::|||      |||
 1233 hrValGlyAsnGlyAlaGluIleAsnAlaThrGluGlyAlaAlaThrLeu 1249

 3830 CCGTTTTCGGGCAATACGGCATCGACA......GGTTCTACATCGGCATC 3873
         :::|||||||||   :::|||      |||||||:::   :::||
 1250 ThrAlaThrGlyAsnThrLeuThrThrGluAlaGlySerSerIleThrSe 1266

 3874 AGCGCGGGCGCGGGTTTTAGCAGCGGCAGCCTTTCAGACGGCATCGGAGG 3923
      |:::::::::::|||         |||      |||:::::::|||   |
 1266 rThrLysGlyGlnValAspLeuLeuAla......GlnAsnGlySerIleA 1281

 3924 CAAAATCCGCCGCCGCGTGCTGCATTACGGCATTCAGGCACGATACCGCG 3973
      ||   |||      |||:::   :::|||                  |||
 1281 laGlySerIleAsnAlaAlaAsnValThrLeuAsnThrThrGlyThrLeu 1297

 3974 CCGGTTTCGGCGGATTCGGCATCGAACCGCACATCGGCGCAACGCGCTAT 4023
         :::   |||   |||     :::   |||||||:::      :::||
 1298 ThrThrValAlaGlySerAspIleLysAlaThrSerGlyThrLeuValIl 1314

 4024 TTCGTCCAAA 4033
      |:::::::|||
 1314 eAsnAlaLys 1317
```

seq_name: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-617-697-4

```
seq_documentation_block:
; Sequence 4, Application US/08617697
; Patent No. 5977336
;  GENERAL INFORMATION:
;    APPLICANT:  Barenkamp, Stephen J
;    TITLE OF INVENTION:  High Molecular Weight Surface Proteins
;    TITLE OF INVENTION:  of No. 5977336-Typeable Haemophilus
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Shoemaker and Mattare, Ltd.
;      STREET:  2001 Jefferson Davis Hwy., 1203 Crystal Plaza
;      STREET:  Bldg. 1
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202-0286
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/617,697
;      FILING DATE:  01-APR-1996
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/302,832
;      FILING DATE:  05-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US PCT/US93/02166
;      FILING DATE:  16-MAR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Berkstresser, Jerry W
;      REGISTRATION NUMBER:  22,651
;      REFERENCE/DOCKET NUMBER:  1038-557
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 415-0810
;      TELEFAX:  (703) 415-0813
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1477 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-617-697-4


alignment_scores:
            Quality:  282.50            Length:   1388
              Ratio:  0.428               Gaps:     61
 Percent Similarity:  47.550   Percent Identity: 19.380

alignment_block:
 US-09-303-518D-649 x US-08-617-697-4   ..

 Align seg 1/1  to: US-08-617-697-4  from: 1  to: 1477

     175 GAAAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAA 224
         :::|||:::|||   ::::::|||:::::::   :::|||::::::
     219 LysAsnGluGlyValIleSerValAsnGlyGlySerIleSerLeu..... 233

     225 CAAAAAAGGGGAGTTGGTCGGCAAATCAATGACAAAAGCC.......... 264
                       |||:::|||::::::::::|||   :::
     234 .............LeuAlaGlyGlnLysIleThrIleSerAspIleIleA 246

     265 ..CCGATGATTGATTTTTCTGTGGTGTCGCGTAACGGCGTGGCGGCATTG 312
           |||   |||   :::|||:::::::::   ::::::   |||:::
     246 snProThrIleThrTyrSerIleAlaAlaProGluAsnGluAlaValAsn 262

     313 GTGGGCGATCAATATATTGTGAGCGTGGCACATAACGGCGGCTATAACAA 362
         :::||||||                 :::      :::||||||      ||
     263 LeuGlyAsp..............IlePheAlaLysGlyGlyAsnIleAs 274
```



```
     |:::   :::|||   :::                    |||   |||   |
 725 nGluAsnMetAsnThrSerLysProLeuProIleArgPheLeuAla...A 741

2051 ACTTCCAAATTAAAGGCGGACAGGCGGTGGTT..................... 2082
     ||:::        ||||||   :::|||
 741 snIleThrAlaThrGlyGlyGlySerValPhePheAspIleTyrAlaAsn 757

2083 ...TCCCGCAATGTTGCCAAAGTGAAA...GGCGATTGGCATTTGAGCAA 2126
        |||   :::   |||:::::|||   ::::::   ::::::|||||
 758 HisSerGlyArgGlyAlaGluLeuLysMetSerGluIleAsnIleSerAs 774

2127 TCACGCCCAAGCAGTTTTTGGTGTCGCACCGCATCAAAGCCACACAATCT 2176
     |   |||:::   ::::::                 :::||||||
 774 nGlyAlaAsnPheThrLeu.....AsnSerHis..... 783

2177 GTACACGTTCGGACTGGACGGGTCTGACAAATTGTGTCGAAAAAACCATT 2226
       :::|||:::|||   :::          |||      :::   ||||||
 784 ..ValArgGlyAspAspAlaPheLysIleAsn...LysAspLeuThrIle 798

2227 ACCGACGATAAAGTGATTGCTTCATTGACTAAGACCGACATCAGCGGCAA 2276
     :::      :::         ||||||   :::|||      :::
 799 AsnAlaThrAsnSerAsnPheSerLeuArgGlnThrLysAspAspPheTy 815

2277 TGTCGATCTTGCCGATCACGCTCATTTAAATCTCACAGGGCTTGCCACAC 2326
                |||   :::|||      :::      :::::::::   |
 815 rAspGlyTyrAlaArgAsnAlaIleAsnSerThrTyrAsnIleSerIleL 832

2327 TCAACGGCAATCTTAGTGCAAATGGCGATACACGTTATACAGTCAGCCAC 2376
     ||:::||||||:::::   :::|||
 832 euGlyGlyAsnValThrLeuGlyGly..... 840

2377 AACGCCACCCAAAACGGCAACCTTAGCCTCGTGGGCAATGCCCAAGCAAC 2426
               ||||||:::::   |||:::::||||||          ||
 841 .........GlnAsnSerSerSerSerIleThrGlyAsn......IleTh 852

2427 ATTTAATCAAGCC.........ACATTAAACGGCAACACATCGGCTTCGG 2467
     |:::::::::|||         ||||||:::::|||:::::   :::
 852 rIleGluLysAlaAlaAsnValThrLeuGluAlaAsnAsnAlaProAsnG 869

2468 GCAATGCTTCATTTAATCTAAGCGACCACGCCGTACAAAACGGCAGTCTG 2517
       :::         |||:::   |||:::::::::::::   |||||||||
 869 lnGln.........AsnIleArgAspArgValIleLysLeuGlySerLeu 882

2518 ACGCTTTCCGGCAACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACGG 2567
       ::::::|||:::   :::   :::   :::::::::   ::::::||
 883 LeuValAsnGlySerLeuSerLeuThrGlyGluAsnAlaAspIleLysGl 899

2568 TAATGTCTCCCTAGCCGATAAGGCAGTATTCCAT.........TTTGAAA 2608
     ||||:::::::::::::::::::|||:::|||            ::::
 899 yAsnLeuThrIleSerGluSerAlaThrPheLysGlyLysThrArgAspT 916

2609 GCAGCCGCTTTACCGGACAAATCAGCGGCGGCAAGGATACGGCATTACAC 2658
     ::   ::::::||||||:::::::::::::::      :::   ::::::
 916 hrLeuAsnIleThrGlyAsnPheThrAsnAsnGlyThrAlaGluIleAsn 932

2659 TTAAAAGACAGCGAATGGACGCTGCCGTCAGGCACGGAATTAGGCAATTT 2708
     :::   ::::::         |||   :::   |||:::   |||:::||
 933 IleThrGlnGlyValValLysLeuGlyAsnValThrAsnAspGlyAspLe 949

2709 AAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCGCCACGATG 2758
     ||||:::   :::
 949 uAsnIleThrThr..... 953

2759 CGGCAGGGGCGCAAACCGGCAGTGCGACAGATGCGCCGCGCCGCCGTTCG 2808
                                                ::::::
 954 .....HisAla 955

2809 CGCCGTTCGCGCCGTTCCCTATTATCCGT....TACACCGCCAACTTCGG 2854
     :::|||::::::||||||:::::::::                ::   ||
 956 LysArgAsnGlnArgSerIleIleGlyGlyAspIleIleAsnLysLysGl 972

2855 TAGAATCCCGTTTCAACACGCTGACGGTAAACGGCAAATTGAACGGTCAG 2904
     |   :::   :::      :::::::::      :::|||:::   :::|
```

```
 972 ySerLeuAsnIleThrAspSerAsnAsnAspAlaGluIleGlnIleGlyG 989

2905 GGAACATTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGACAAATT 2954
     ||||||||           |||:::|||   :::    |||||||||::
 989 lyAsnIleSerGlnLysGluGlyAsnLeuThrIle.SerSerAspLysIl 1005

2955 GAAGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCGGTCAACAATA 3004
     ::::::::::
1005 eAsnIleThr..... 1008

3005 CCGGCAACGAACCTGCAAGCCTCGAACAATTGACGGTAGTGGAAGGAAAA 3054
                        :::|||:::|||:::   :::|||
1009 .....LysGlnIleThrIleLysLysGlyIle 1017

3055 GACAACAAACCGCTGTCC...........GAAAACCTTAATTTCACCCT 3092
     |||::::::      |||           :::|||   |||:::|||::
1018 AspGlyGluAspSerSerSerAspAlaThrSerAsnAlaAsnLeuThrIl 1034

3093 GCAAAACGAACACGTCGAT.....................GCCGGCGCGT 3121
     ::::::::::::::::                        :::|||
1034 eLysThrLysGluLeuLysLeuThrGluAspLeuSerIleSerGlyPheA 1051

3122 GGCGTTACCAACTCATCCGCAAAGACGGC...GAGTTCCGCCTGCATAAT 3168
       :::   ::::::      |||||||||   ::::::   :::   |||
1051 snLysAlaGluIleThrAlaLysAspGlyArgAspLeuThrIleGlyAsn 1067

3169 CCGGTCAAAGAACAAGAGCTTTCCGACAAACTCGGCAAGGCAGAAGCCAA 3218
                          :::|||      :::   ||||||||
1068 Ser.................AsnAspGlyAsnSerGlyAlaGluAla.. 1077

3219 AAAACAGGCGGAAAAAGACAACGCGCAAAGCCTTGACGCGCTGATTGCGG 3268
      |||   :::      :::|||::::::      |||:::   |||:::|
1078 .LysThrValThrPheAsnAsnValLys......AspSerLysIleSerA 1092

3269 CCGGGCGCGATGCCGTCGAAAAGACAGAAAGCGTTGCCGAACCGGCCCGG 3318
     ||            |||      :::::::::|||          :::
1092 laAspGlyHisAsnValThrLeuAsnSerLysValLysThrSerSerSer 1108

3319 CAGGCAGGCGGGGAAAATGTCGGCATTATGCAGGCGGAGGAAGAGAAAAA 3368
     ::::::|||   |||:::
1109 AsnGlyGlyArgGluSer..... 1114

3369 ACGGGTGCAGGCGGATAAAGACACCGCCTTGGCGAAACAGCGCGAAGCGG 3418
            ::::::|||:::|||||||:::|||
1115 .......AsnSerAspAsnAspThrGlyLeu..... 1122

3419 AAACCCGGCCGGCTACCACCGCCTTCCCCCGCGCCCGCCGCGCCCGCCGG 3468
                     |||         :::      ::::::::::::
1123 .................ThrIleThrAlaLysAsnValGluValAsnLys 1133

3469 GATTTGCCGCAACTGCAACCCCAACCGCAGCCCCAACCGCAGCGCGACCT 3518
     |||:::   :::|||
1134 AspIleThrSerLeu..... 1138

3519 GATCAGCCGTTATGCCAATAGCGGTTTGAGTGAATTTTCCGCCACGCTCA 3568
            :::   :::|||         ||||||      :::||| :::
1139 .......LysThrValAsnIleThrAlaSerGluLysValThrThr.Thr 1152

3569 ACAGCGTTTTCGCCGTACAGGACGAATTAGACCGCGTATTTGCCGAAGAC 3618
     ::::::   :::         ||||||      |||   :::
1153 AlaGlySerThrIleAsnAlaThrAsnGlyLysAlaSerIleThrThrLy 1169

3619 CGCCGCAACGCCGTTTGGACAAGCGGCATCCGGGACACCAAACACTACCG 3668
      ::::::         |||      :::::::::|||   ::::::::::
1169 sThrGlyAspIleSerGlyThrIleSerGlyAsnThrValSerValSerA 1186

3669 TTCGC...AAGATTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCAAA 3715
     ::      :::::::::   :::::::::|||      :::   :::
1186 laThrValAspLeuThrThrLysSerGlySerLysIleGluAlaLysSer 1202

3716 TCGGTATGCAGAAAAACCTCGGCAGCGGGCGCGTCGGCATCCTGTTTTCG 3765
     :::   ::::::   ||||||||||:::|||:::   ::::::
1203 GlyGluAlaAsnValThrSerAlaThrGlyThrIleGlyGly..... 1216
```

```
    3766 CACAACCGGACCGAAAACACCTTCGACGACGGCATCGGCA..... 3805
         |||   :::   :::|||   ::::::||||||:::|||
    1217 .ThrIleSerGlyAsnThrValAsnValThrAlaAsnAlaGlyAspLeuT 1233

    3806 .........................ACTCGGCACGGCTTGCCCACGGCG 3829
                                   :::   :::|||      |||
    1233 hrValGlyAsnGlyAlaGluIleAsnAlaThrGluGlyAlaAlaThrLeu 1249

    3830 CCGTTTTCGGGCAATACGGCATCGACA......GGTTCTACATCGGCATC 3873
              :::||||||||||   :::|||      ||||||:::   ::>||
    1250 ThrAlaThrGlyAsnThrLeuThrThrGluAlaGlySerSerIleThrSe 1266

    3874 AGCGCGGGCGCGGGTTTTAGCAGCGGCAGCCTTTCAGACGGCATCGGAGG 3923
         |:::::::::::|||         |||      |||:::::::|||   |
    1266 rThrLysGlyGlnValAspLeuLeuAla......GlnAsnGlySerIleA 1281

    3924 CAAAATCCGCCGCCGCGTGCTGCATTACGGCATTCAGGCACGATACCGCG 3973
         ||   |||      |||:::   :::|||                .|||
    1281 laGlySerIleAsnAlaAlaAsnValThrLeuAsnThrThrGlyThrLeu 1297

    3974 CCGGTTTCGGCGGATTCGGCATCGAACCGCACATCGGCGCAACGCGCTAT 4023
            :::   |||   |||      :::   ||||||:::      :::||
    1298 ThrThrValAlaGlySerAspIleLysAlaThrSerGlyThrLeuValIl 1314

    4024 TTCGTCCAAA 4033
         |:::::|||
    1314 eAsnAlaLys 1317
```

seq_name: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-719-641-4

```
seq_documentation_block:
; Sequence 4, Application US/08719641
; Patent No. 6218141
;  GENERAL INFORMATION:
;    APPLICANT:  Barenkamp, Stephen J
;    TITLE OF INVENTION:  High Molecular Weight Surface Proteins
;    TITLE OF INVENTION:  of No. 6218141-Typeable Haemophilus
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Shoemaker and Mattare, Ltd.
;      STREET:  2001 Jefferson Davis Hwy., 1203 Crystal Plaza
;      STREET:  Bldg. 1
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202-0286
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/719,641
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/302,832
;      FILING DATE:  16-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US PCT/US93/02166
;      FILING DATE:  16-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9205704.1
;      FILING DATE:  16-MAR-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Berkstresser, Jerry W
;      REGISTRATION NUMBER:  22,651
;      REFERENCE/DOCKET NUMBER:  1038-625
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 415-0810
;      TELEFAX:  (703) 415-0813
;  INFORMATION FOR SEQ ID NO:  4:
```

```
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1477 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-719-641-4


alignment_scores:
            Quality:  282.50              Length:   1388
              Ratio:  0.428                 Gaps:     61
 Percent Similarity:  47.550   Percent Identity: 19.380

alignment_block:
 US-09-303-518D-649 x US-08-719-641-4   ..

 Align seg 1/1  to: US-08-719-641-4  from: 1  to: 1477

     175 GAAAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAA 224
         :::|||:::|||   ::::::|||::::::   :::|||::::::
     219 LysAsnGluGlyValIleSerValAsnGlyGlySerIleSerLeu..... 233

     225 CAAAAAAGGGGAGTTGGTCGGCAAATCAATGACAAAAGCC.......... 264
                     |||:::|||:::::::::|||   :::
     234 .............LeuAlaGlyGlnLysIleThrIleSerAspIleIleA 246

     265 ..CCGATGATTGATTTTTCTGTGGTGTCGCGTAACGGCGTGGCGGCATTG 312
           |||   |||   :::|||:::::::::   ::::::   |||:::
     246 snProThrIleThrTyrSerIleAlaAlaProGluAsnGluAlaValAsn 262

     313 GTGGGCGATCAATATATTGTGAGCGTGGCACATAACGGCGGCTATAACAA 362
         :::||||||                :::      :::||||||      ||
     263 LeuGlyAsp..............IlePheAlaLysGlyGlyAsnIleAs 274

     363 CGTT..........................GATTTTGGTGCGGAAGGAA 385
         ||||                         ::::::|||::::::
     274 nValArgAlaAlaThrIleArgAsnGlnGlyLysLeuSerAlaAspServ 291

     386 GAAATCCCGATCAACATCGTTTTACTTATAAAATTGTGAAACGGAATAAT 435
           :::   |||:::                        :::|||:::::::::
     291 alSerLysAspLysSerGlyAsnIleValLeuSerAlaLysGluGlyGlu 307

     436 TATAAAGCAGGG..............ACTAAAGGCCATCCTTATGGCGG 470
            :::   |||                  :::   :::      |||||
     308 AlaGluIleGlyGlyValIleSerAlaGlnAsnGlnGlnAlaLysGlyGl 324

     471 CGATTATCATATGCCGCGTTTGCATAAATTTGTCACAGATGCAGAACCTG 520
         |      :::                  :::      :::      :
     324 yLysLeuMetIleThrGlyAspLysValThrLeuLysThrGlyAlaValI 341

     521 TTGAAATGACCAGTTATATGGATGGGCGGAAATATATCGATCAAAATAAT 570
         ::::::::::::::         |||:::   |||:::      ::::::
     341 leAspLeuSerGlyLysGluGlyGlyGluThrTyrLeuGlyGlyAspGlu 357

     571 TACCCTGACCGTGTTCGTATTGGGGCAGGCAGGCAATATTGGCGATCTGA 620
                              |||   |||::::::      :::
     358 Arg.................GlyGluGlyLysAsnGlyIleGlnLeuAl 368

     621 TGAAGATGAGCCCAATAACCGCGAAAGTTCATATCATATTGCAAGT.... 666
          :::          ::::::   |||:::   ::::::::::::
     368 aLysLysThrSerLeuGluLysGlySerThrIleAsnValSerGlyLysG 385

     667 .............GCGTATTCTTGG............CTCGTTGGTGGC 690
                     |||      |||            |||:::   |||
     385 luLysGlyGlyPheAlaIleValTrpGlyAspIleAlaLeuIleAspGly 401

     691 AATACCTTTGCACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAG 740
         |||      ||||||   |||||||||
     402 AsnIleAsnAlaGln...GlySerGly..................... 409

     741 TGAAAAAATTAAACATAGCCCATATGGTTTTTTACCAACAGGAGGCTCAT 790
                                  :::      |||||||||
     410 .......................AspIleAlaLysThrGlyGlyPheV 418
```

```
 791 TTGGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAG 840
418 alGluThrSerGlyHisAspLeuPheIleLysAsp..............  429

 841 TGGTTAATTAATGGGGTATTGCAAACGGGCAACCCCTATATAGGAAAAAG 890
430 .........AsnAlaIleValAspAla.......................  435

 891 CAATGGCTTCCAGCTGGTTCGTAAAGATTGGTTCTATGATGAAATCTTTG 940
436 .....................LysGluTrpLeuLeuAsp..........  441

 941 CTGGAGATACCCATTCAGTATTCTACGAACCACGTCAAAATGGGAAATAC 990
442 ..PheAspAsnValSerIleAsnAlaGluAspProLeuPheAsnAsnThr 457

 991 TCTTTTAACGACGAT...AATAATGGCACAGGAAAAATCAATGCCAAACA 1037
458 GlyIleAsnAspGluPheProThrGlyThrGlyGluAlaSerAspProLy 474

1038 TGAACACAATTCTCTGCCTAATAGATTAAAAACACGAACCGTTCAATTGT 1087
474 sLysAsnSerGluLeuLysThrThrLeuThrAsnThrThrIleSerAsnT 491

1088 TT.....................AATGTTTCTTTATCCGAGACAGCAAGA 1116
491 yrLeuLysAsnAlaTrpThrMetAsnIleThrAlaSerArgLysLeuThr 507

1117 GAACCTGTTTATCATGCTGCAGGTGGTGTCAACAGTTATCGACCCAGACT 1166
508 ..........................ValAsnSer.........SerIl 512

1167 GAATAATGGAGAAAATATTTCCTTTATTGACGAAGGAAAAGGCGAATTGA 1216
512 eAsnIleGlySerAsnSerHisLeuIleLeuHisSerLysGly.......  526

1217 TACTTACCAGCAACATCAATCAAGGTGCTGGAGGATTATATTTCCAAGGA 1266
527 ....................GlnArgGlyGlyGlyValGlnIleAspGly 536

1267 GATTTTACGGTCTCGCCTGAAAATAACGAAACTTGGCAAGGCGCGGGCGT 1316
537 AspIleThrSerLysGlyGlyAsnLeuThrIleTyrSerGlyGlyTrpVa 553

1317 TCATATCAGTGAAGACAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAA 1366
553 lAspValHisLysAsnIleThrLeu.........................  561

1367 ACGACCGCCTGTCCAAAATCGGCAAAGGCACGCTGCACGTTCAAGCCAAA 1416
561 ..................................................  561

1417 GGGGAAAACCAAGGCTCGATCAGCGTGGGCGACGGTACAGTCATTTTGGA 1466
562 ......AspGlnGlyPheLeuAsnIleThrAlaAlaSerValAlaPhe..  575

1467 TCAGCAGGCAGACGATAAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCT 1516
576 .GluGlyGlyAsnAsnLysAlaArgAspAlaAlaAsnAlaLysIle....  590

1517 TGGTCAGCGGCAGGGGTACGGTGCAACTGAATGCCGATAATCAGTTCAAC 1566
591 .....ValAlaGlnGlyThrValThrIleThrGlyGluGlyLys......  603

1567 CCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAACGG 1616
604 ..................AspPheArgAlaAsnAsnValSerLeuAsnGl 614

1617 GCATTCGCTTTCGTTCCACCGTATTCAAAATACCGATGAAGGGGCGATGA 1666
614 y............................ThrGlyLysGlyLeuAsnI 621
```

```
1667 TTGTC............AACCACAATCAAGACAAAGAATCCACCGTTACC 1704
 621 leIleSerSerValAsnAsnLeuThrHisAsnLeuSerGlyThrIleAsn 637

1705 ATTACAGGCAATAAAGATATTGCTACAACCGGCAATAACAACAGCTTGGA 1754
 638 IleSerGlyAsnIleThrIleAsnGlnThr..................... 647

1755 TAGCAAAAAAGAAATTGCCTACAACGGTTGGTTTGGCGAGAAAGATACG. 1803
 648 .ThrArgLysAsnThrSerTyr......TrpGlnThrSerHisAspSerH 662

1804 .......................ACCAAAACGAACGGGCGGCTCAACCTT 1830
 662 isTrpAsnValSerAlaLeuAsnLeuGluThrGlyAlaAsnPheThrPhe 678

1831 GTTTACCAGCCCGCCGCAGAAGACCGCACCCTGCTGCTT........... 1869
 679 IleLysTyrIleSerSerAsnSerLysGlyLeuThrThrGlnTyrArgSe 695

1870 .TCCGGCGGAACAAATTTAAACGGCAACATCACGCAAACAAACGGCAAAC 1918
 695 rSerAlaGlyValAsnPheAsnGly............ValAsnGlyAsnM 708

1919 TGTTTTTCAGCGGCAGACCAACACCGCACGCCTACAATCATTTAAACGAC 1968
 708 etSerPhe.......................................... 710

1969 CATTGGTCGCAAAAAGAGGGCATTCCTCGCGGGGAAATCGTGTGGGACAA 2018
 711 ......AsnLeuLysGluGlyAlaLysValAsnPheLysLeuLysProAs 725

2019 CGACTGGATCAACCGCACA..................TTTAAAGCGGAAA 2050
 725 nGluAsnMetAsnThrSerLysProLeuProIleArgPheLeuAla...A 741

2051 ACTTCCAAATTAAAGGCGGACAGGCGGTGGTT................... 2082
 741 snIleThrAlaThrGlyGlyGlySerValPhePheAspIleTyrAlaAsn 757

2083 ...TCCCGCAATGTTGCCAAAGTGAAA...GGCGATTGGCATTTGAGCAA 2126
 758 HisSerGlyArgGlyAlaGluLeuLysMetSerGluIleAsnIleSerAs 774

2127 TCACGCCCAAGCAGTTTTGGTGTCGCACCGCATCAAAGCCACACAATCT 2176
 774 nGlyAlaAsnPheThrLeu..............AsnSerHis....... 783

2177 GTACACGTTCGGACTGGACGGGTCTGACAAATTGTGTCGAAAAAACCATT 2226
 784 ..ValArgGlyAspAspAlaPheLysIleAsn...LysAspLeuThrIle 798

2227 ACCGACGATAAAGTGATTGCTTCATTGACTAAGACCGACATCAGCGGCAA 2276
 799 AsnAlaThrAsnSerAsnPheSerLeuArgGlnThrLysAspAspPheTy 815

2277 TGTCGATCTTGCCGATCACGCTCATTTAAATCTCACAGGGCTTGCCACAC 2326
 815 rAspGlyTyrAlaArgAsnAlaIleAsnSerThrTyrAsnIleSerIleL 832

2327 TCAACGGCAATCTTAGTGCAAATGGCGATACACGTTATACAGTCAGCCAC 2376
 832 euGlyGlyAsnValThrLeuGlyGly.....................  840

2377 AACGCCACCCAAAACGGCAACCTTAGCCTCGTGGGCAATGCCCAAGCAAC 2426
 841 .........GlnAsnSerSerSerSerIleThrGlyAsn......IleTh 852

2427 ATTTAATCAAGCC.........ACATTAAACGGCAACACATCGGCTTCGG 2467
 852 rIleGluLysAlaAlaAsnValThrLeuGluAlaAsnAsnAlaProAsnG 869

2468 GCAATGCTTCATTTAATCTAAGCGACCACGCCGTACAAAACGGCAGTCTG 2517
```



```
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Shoemaker and Mattare, Ltd
;      STREET:  2001 Jefferson Davis Hwy., 1203 Crystal Plaza
;      STREET:  Bldg. 1
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202-0286
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/038,682
;      FILING DATE:  16-MAR-1993
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BERKSTRESSER, JERRY W
;      REGISTRATION NUMBER:  22,651
;      REFERENCE/DOCKET NUMBER:  1038-293
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 415-0810
;      TELEFAX:  (703) 415-0813
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1536 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-038-682-2


alignment_scores:
            Quality:  272.50              Length:   1182
              Ratio:  0.452                 Gaps:     60
 Percent Similarity:  51.015   Percent Identity: 20.897

alignment_block:
 US-09-303-518D-649 x US-08-038-682-2   ..

 Align seg 1/1  to: US-08-038-682-2  from: 1  to: 1536

     127 GGACACACTTATTTCGGCATCAACTACCAATACTATCGCGACTTTGCCGA 176
         |||:::||||||::|||                          :::::
     349 GlyGluThrTyrLeuGly.........................GlyAs 356

     177 AAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACA 226
         :::::::|||:::          |||   :::   |||::::::   :::|
     356 pGluArgGlyGlu.........GlyLysAsnGlyIleGlnLeuAlaLysL 370

     227 AAAAAGGGGAGTTGGTCGGCAAATCAATGACAAAAGCCCCGATGATTGAT 276
         ||                     |||:::   |||:::      |||:::
     370 ys.................ThrSerLeuGluLysGlySerThrIleAsn 380

     277 TTTTCTGTGGTGTCGCGTAACGGCGTGGCGGCATTGGTGGGCGATCAATA 326
            |||      ::::::::::|||   |||   :::   ||||||
     381 ValSerGlyLysGluLysGlyGlyArgAlaIleValTrpGlyAspIleAl 397

     327 TATTGTG............................AGCGTGGCACATA 346
          ::::::                           ::::::|||   :
     397 aLeuIleAspGlyAsnIleAsnAlaGlnGlySerGlyAspIleAlaLysT 414

     347 ACGGCGGC..................................... 354
         ::||||||
     414 hrGlyGlyPheValGluThrSerGlyHisAspLeuPheIleLysAspAsn 430

     355 ..............................TATAACAACGTTGATTT 371
                                      ::::::||||||:::::
     431 AlaIleValAspAlaLysGluTrpLeuLeuAspPheAspAsnValSerIl 447

     372 TGGTGCGGAA......GGAAGAAATCCCGATCAACATCGTTTTACTTAT. 414

         ::::||||||      ||||||:::      ::::::         |||
     447 eAsnAlaGluThrAlaGlyArgSerAsnThrSerGluAspAspGluTyrT 464

     415 ...................AAAATTGTGAAACGGAATAATTATAAAGCA 444
                               :::      |||||||||:::   |||:::
     464 hrGlySerGlyAsnSerAlaSerThrProLysArgAsnLysGluLysThr 480

     445 GGGACTAAAGGCCATCCTTATGGCGGCGATTATCATATGCCGCGTTTGCA 494
             |||            :::   :::|||:::         :::   :
     481 .ThrLeuThrAsnThrThrLeuGluSerIleLeuLysLysGlyThrPheV 497

     495 TAAATTTGTCACAGATGCAGAACCTGTTGAAATGACCAGTTATATGGATG 544
         ::|||::::::      |||:::                  |||::::::
     497 alAsnIleThrAlaAsnGlnArg.................IleTyrVal 507

     545 GGCGGAAATATATCGATCAAAATAATTACCCTGACCGTGTTCGTATTGGG 594
         :::::::::|||:::::::::      :::||||||:::::      ||
     508 AsnSerSerIleAsnLeuSerAsnGlySerLeuThrLeuTrpSerGluGl 524

     595 GCAGGCAGGC..........................AATATTGGCGATC 617
         |:::::::|||                         :::   ||||||
     524 yArgSerGlyGlyGlyValGluIleAsnAsnAspIleThrThrGlyAspA 541

     618 TGATGAAGATGAGCCCAATAACC.................GCGAAAGTT 649
             :::   :::   :::|||                  :::   |||
     541 spThrArgGlyAlaAsnLeuThrIleTyrSerGlyGlyTrpValAspVal 557

     650 CATATCATATTGCAAGTGCGTATTCTTGGCTCGTTGGTGGCAATACCTTT 699
         |||      :::::::::         |||::::::  :::   |||
     558 HisLysAsnIleSerLeuGlyAlaGlnGlyAsnIle..AsnIleThrAla 573

     700 GCACAAAAT..............GGATCA..............GGTGG 719
            |||:::               ||||||               |||
     574 LysGlnAspIleAlaPheGluLysGlySerAsnGlnValIleThrGlyGl 590

     720 TGGCACAGTCAACTTAGGTAGTGAAAAAATTAAACATAGCCCATATGGTT 769
          ||||||::::::   |||::::::|||                  ||||
     590 nGlyThrIleThrSerGlyAsnGlnLys.................GlyP 601

     770 TTTTACCAACAGGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATC 819
         ||      ::::::   |||::::::   ::::::::   :::   :::
     601 heArgPheAsnAsnValSerLeuAsnGlyThrGlySerGlyLeuGlnPhe 617

     820 TATGATGCCCAAAAGCAAAAGTGGTTAATTAATGGGGTATTGCAAACGGG 869
                :::   :::|||:::   |||::::::   ::::::
     618 ThrThrLysArgThrAsnLysTyrAlaIleThrAsnLysPheGluGlyTh 634

     870 CAACCCCTATATAGGAAAAAGCAAT...GGCTTCCAGCTGGTTCGTAAAG 916
                      ||||||   |||   ::::::   |||   :::::::
     634 rLeuAsnIleSerGlyLysValAsnIleSerMetValLeuProLysAsnG 651

     917 ATTGGTTCTATGATGAAATCTTTGCTGGAGATACCCAT......TCAGTA 960
         ::      ||||||:::   |||   |||   |||:::         :::
     651 luSerGlyTyrAspLys...PheLysGlyArgThrTyrTrpAsnLeuThr 666

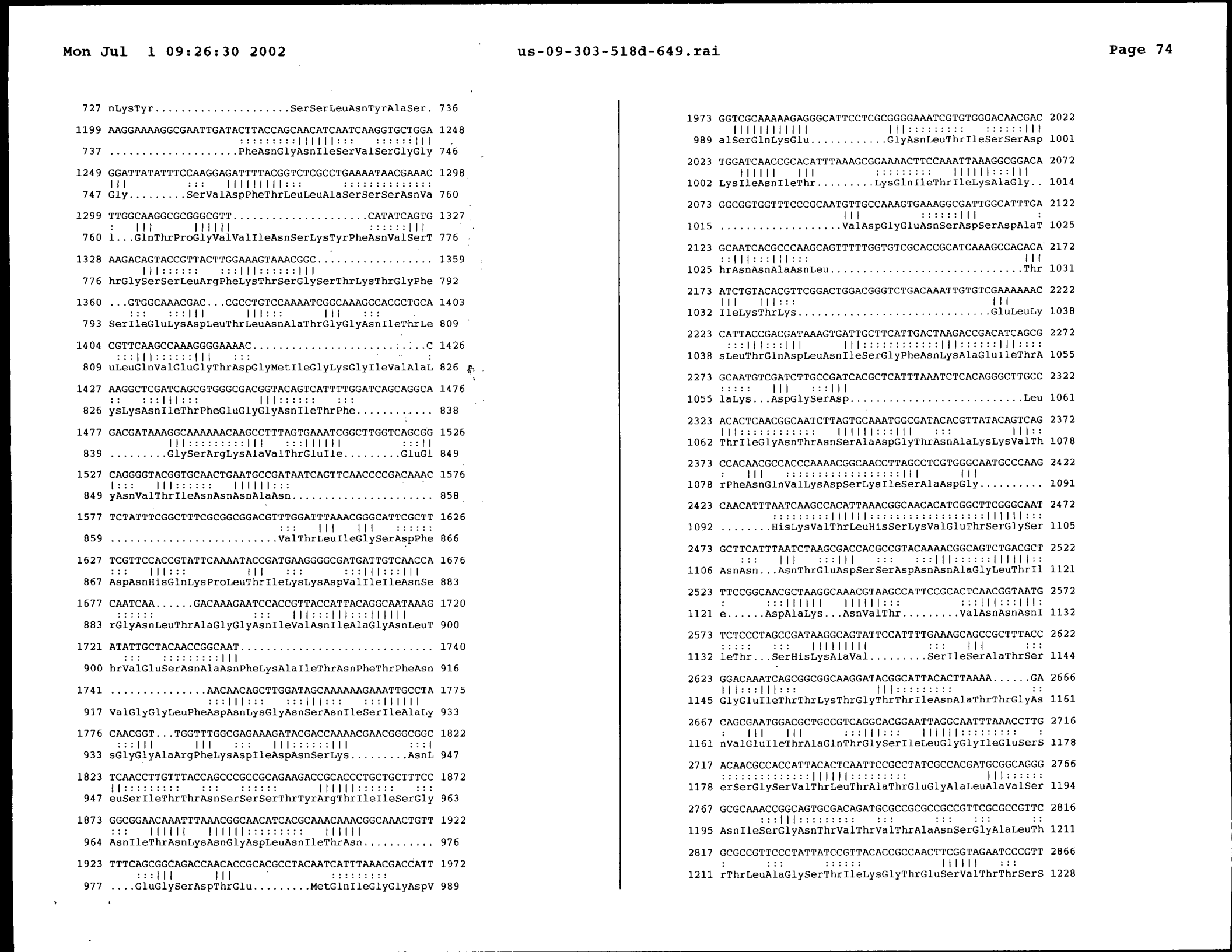
     961 TTCTACGAACCACGTCAAAATGGGAAATACTCTTTTAACGACGATAATAA 1010
             :::      :::::::|||:::::::::::::::   |||::::::
     667 SerLeuAsnValSerGluSerGlyGluPheAsnLeuThrIleAspSerAr 683

    1011 TGGCACA.........GGAAAAATCAATGCCAAACATGAACACAATTCTC 1051
         :|||:::         |||   ::::::      ::::::   |||::::
     683 gGlySerAspSerAlaGlyThrLeuThrGlnProTyrAsnLeuAsnGlyI 700

    1052 TGCCT...AATAGATTAAAAACACGAACCGTTCAATTGTTTAATGTTTCT 1098
         ::      |||:::      |||   :::|||:::      |||:::
     700 leSerPheAsnLysAspThrThrPheAsnValGluArg...AsnAlaArg 715

    1099 TTATCCGAGACAGCAAGAGAACCTGTTTATCATGCTGCAGGTGGTGTCAA 1148
         ::::::         :::   |||:::                |||:::||
     716 ValAsnPheAspIleLysAlaProIle..............GlyIleAs 727

    1149 CAGTTATCGACCCAGACTGAATAATGGAGAAAATATTTCCTTTATTGACG 1198
         |:::|||                      ::::::::::::::   :::
```



```
     2867 TCAACACGCTGACGGTAAACGGCAAATTGAACGGTCAGGGAACA..... 2910
          ::::::      :::::::|||   :::::::|||   ||||||
     1228 erGlnSerGlyAspIleGlyGlyThrIleSerGly...GlyThrValGlu 1243

     2911 TTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGACAAATTGAAGCT 2960
            :::      :::::::|||            ::::::|||:::|||
     1244 ValLysAlaThrGluSerLeuThrThrGlnSerAsnSerLysIleLys.. 1259

     2961 GGCGGAAAGTTCCGAAGGCACTTACACCTTGGCGGTC............. 2997
          |||   ::::::|||::::::   |||   |||:::
     1260 .AlaThrThrGlyGluAlaAsnValThrSerAlaThrGlyThrIleGlyG 1276

     2998 ..........AACAATACCGGCAACGAACCTGCAAGCCTCGAACAATTG 3036
                  :::|||||||   |||      |||:::      :::|||
     1276 lyThrIleSerGlyAsnThrValAsnValThrAlaAsnAlaGlyAspLeu 1292

     3037 ACGGTA.......................GTGGAAGGAAAAGACAACAA 3062
          ||||||                      :::||||||      :::
     1293 ThrValGlyAsnGlyAlaGluIleAsnAlaThrGluGlyAlaAlaThrLe 1309

     3063 ACCGCTGTCCGAAAACCTTAATTTCACCCTGCAAAACGAACACGTCGATG 3112
               |||::::::         |||      ::::::|||:::   :
     1309 uThrThrSerSerGlyLysLeuThrThrGluAlaSerSerHisIleThrS 1326

     3113 CCGGCGCGTGGCGTTACCAACTCATCCGCAAAGACGGCGAGTTCCGCCTG 3162
          :: |||:::|||   :::|||   ||||||:::   :::|||   |||
     1326 er.AlaLysGlyGlnValAsnLeuSerAlaGlnAspGlySerValAlaGl 1342

     3163 CATAATCCGGTCAAAGAACAAGAGCTTTCCGACAAACTCGGCAAGGCAGA 3212
               |||::::::   |||   ::::::   |||:::::::::
     1342 ySerIleAsnAlaAlaAsnValThrLeuAsnThrThrGlyThrLeuThrT 1359

     3213 AGCCAAAAAACAGGCGGAAAAAGACAACGCGCAAAGCCTTG 3253
               |||:::   :::   :::::::|||      :::|||
     1359 hrValLysGlySerAsnIleAsnAlaThrSerGlyThrLeu 1372
```

seq_name: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-302-832-2

```
seq_documentation_block:
; Sequence 2, Application US/08302832
; Patent No. 5603938
;  GENERAL INFORMATION:
;    APPLICANT:  Barenkamp, Stephen J
;    TITLE OF INVENTION:  High Molecular Weight Surface Proteins
;    TITLE OF INVENTION:  of No. 5603938-Typeable Haemophilus
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Shoemaker and Mattare, Ltd.
;      STREET:  2001 Jefferson Davis Hwy., 1203 Crystal Plaza
;      STREET:  Bldg. 1
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202-0286
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/302,832
;      FILING DATE:  16-SEP-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9205704.1
;      FILING DATE:  16-MAR-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US pct/us93/02166
;      FILING DATE:  16-MAR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Berkstresser, Jerry W
;      REGISTRATION NUMBER:  22,651
;      REFERENCE/DOCKET NUMBER:  1038-404
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 415-0810
;      TELEFAX:  (703) 415-0813
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1536 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-302-832-2


alignment_scores:
            Quality:  272.50              Length:   1182
              Ratio:  0.452                 Gaps:     60
 Percent Similarity:  51.015   Percent Identity: 20.897

alignment_block:
 US-09-303-518D-649 x US-08-302-832-2   ..

 Align seg 1/1  to: US-08-302-832-2  from: 1  to: 1536

      127 GGACACACTTATTTCGGCATCAACTACCAATACTATCGCGACTTTGCCGA 176
          |||:::||||||:::|||                              :::::
      349 GlyGluThrTyrLeuGly.........................GlyAs 356

      177 AAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACA 226
          ::::::::|||:::          |||   :::   |||::::::   :::|
      356 pGluArgGlyGlu.........GlyLysAsnGlyIleGlnLeuAlaLysL 370

      227 AAAAAGGGGAGTTGGTCGGCAAATCAATGACAAAAGCCCCGATGATTGAT 276
          ||                        |||:::  |||:::      |||:::
      370 ys..................ThrSerLeuGluLysGlySerThrIleAsn 380

      277 TTTTCTGTGGTGTCGCGTAACGGCGTGGCGGCATTGGTGGGCGATCAATA 326
            |||       ::::::::::|||   |||   :::   ||||||
      381 ValSerGlyLysGluLysGlyGlyArgAlaIleValTrpGlyAspIleAl 397

      327 TATTGTG..............................AGCGTGGCACATA 346
           ::::::                            ::::::|||   :
      397 aLeuIleAspGlyAsnIleAsnAlaGlnGlySerGlyAspIleAlaLysT 414

      347 ACGGCGGC.......................................... 354
          ::||||||
      414 hrGlyGlyPheValGluThrSerGlyHisAspLeuPheIleLysAspAsn 430

      355 ................................TATAACAACGTTGATTT 371
                                          ::::::||||||:::::
      431 AlaIleValAspAlaLysGluTrpLeuLeuAspPheAspAsnValSerIl 447

      372 TGGTGCGGAA......GGAAGAAATCCCGATCAACATCGTTTTACTTAT. 414
          ::::||||||      ||||||:::      ::::::          |||
      447 eAsnAlaGluThrAlaGlyArgSerAsnThrSerGluAspAspGluTyrT 464

      415 ...................AAAATTGTGAAACGGAATAATTATAAAGCA 444
                             :::      |||||||||:::   |||:::
      464 hrGlySerGlyAsnSerAlaSerThrProLysArgAsnLysGluLysThr 480

      445 GGGACTAAAGGCCATCCTTATGGCGGCGATTATCATATGCCGCGTTTGCA 494
            |||          :::   :::|||:::      :::   :
      481 .ThrLeuThrAsnThrThrLeuGluSerIleLeuLysLysGlyThrPheV 497

      495 TAAATTTGTCACAGATGCAGAACCTGTTGAAATGACCAGTTATATGGATG 544
          ::|||::::::      |||:::                  |||::::::
      497 alAsnIleThrAlaAsnGlnArg.................IleTyrVal 507

      545 GGCGGAAATATATCGATCAAAATAATTACCCTGACCGTGTTCGTATTGGG 594
          :::::::::|||::::::::::      :::||||||::::::      ||
      508 AsnSerSerIleAsnLeuSerAsnGlySerLeuThrLeuTrpSerGluGl 524

      595 GCAGGCAGGC.........................AATATTGGCGATC 617
          |::::::|||                         :::   ||||||
```



```
2223 CATTACCGACGATAAAGTGATTGCTTCATTGACTAAGACCGACATCAGCG 2272
1038 sLeuThrGlnAspLeuAsnIleSerGlyPheAsnLysAlaGluIleThrA 1055

2273 GCAATGTCGATCTTGCCGATCACGCTCATTTAAATCTCACAGGGCTTGCC 2322
1055 laLys...AspGlySerAsp.....Leu 1061

2323 ACACTCAACGGCAATCTTAGTGCAAATGGCGATACACGTTATACAGTCAG 2372
1062 ThrIleGlyAsnThrAsnSerAlaAspGlyThrAsnAlaLysLysValTh 1078

2373 CCACAACGCCACCCAAAACGGCAACCTTAGCCTCGTGGGCAATGCCCAAG 2422
1078 rPheAsnGlnValLysAspSerLysIleSerAlaAspGly..... 1091

2423 CAACATTTAATCAAGCCACATTAAACGGCAACACATCGGCTTCGGGCAAT 2472
1092 ........HisLysValThrLeuHisSerLysValGluThrSerGlySer 1105

2473 GCTTCATTTAATCTAAGCGACCACGCCGTACAAAACGGCAGTCTGACGCT 2522
1106 AsnAsn...AsnThrGluAspSerSerAspAsnAsnAlaGlyLeuThrIl 1121

2523 TTCCGGCAACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACGGTAATG 2572
1121 e......AspAlaLys...AsnValThr.....ValAsnAsnAsnI 1132

2573 TCTCCCTAGCCGATAAGGCAGTATTCCATTTTGAAAGCAGCCGCTTTACC 2622
1132 leThr...SerHisLysAlaVal.....SerIleSerAlaThrSer 1144

2623 GGACAAATCAGCGGCGGCAAGGATACGGCATTACACTTAAAA......GA 2666
1145 GlyGluIleThrThrLysThrGlyThrThrIleAsnAlaThrThrGlyAs 1161

2667 CAGCGAATGGACGCTGCCGTCAGGCACGGAATTAGGCAATTTAAACCTTG 2716
1161 nValGluIleThrAlaGlnThrGlySerIleLeuGlyGlyIleGluSerS 1178

2717 ACAACGCCACCATTACACTCAATTCCGCCTATCGCCACGATGCGGCAGGG 2766
1178 erSerGlySerValThrLeuThrAlaThrGluGlyAlaLeuAlaValSer 1194

2767 GCGCAAACCGGCAGTGCGACAGATGCGCCGCGCCGCCGTTCGCGCCGTTC 2816
1195 AsnIleSerGlyAsnThrValThrValThrAlaAsnSerGlyAlaLeuTh 1211

2817 GCGCCGTTCCCTATTATCCGTTACACCGCCAACTTCGGTAGAATCCCGTT 2866
1211 rThrLeuAlaGlySerThrIleLysGlyThrGluSerValThrThrSerS 1228

2867 TCAACACGCTGACGGTAAACGGCAAATTGAACGGTCAGGGAACA..... 2910
1228 erGlnSerGlyAspIleGlyGlyThrIleSerGly...GlyThrValGlu 1243

2911 TTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGACAAATTGAAGCT 2960
1244 ValLysAlaThrGluSerLeuThrThrGlnSerAsnSerLysIleLys.. 1259

2961 GGCGGAAAGTTCCGAAGGCACTTACACCTTGGCGGTC..... 2997
1260 .AlaThrThrGlyGluAlaAsnValThrSerAlaThrGlyThrIleGlyG 1276

2998 ..........AACAATACCGGCAACGAACCTGCAAGCCTCGAACAATTG 3036
1276 lyThrIleSerGlyAsnThrValAsnValThrAlaAsnAlaGlyAspLeu 1292

3037 ACGGTA.....GTGGAAGGAAAAGACAACAA 3062
1293 ThrValGlyAsnGlyAlaGluIleAsnAlaThrGluGlyAlaAlaThrLe 1309

3063 ACCGCTGTCCGAAAACCTTAATTTCACCCTGCAAAACGAACACGTCGATG 3112
```

```
1309 uThrThrSerSerGlyLysLeuThrThrGluAlaSerSerHisIleThrS 1326

3113 CCGGCGCGTGGCGTTACCAACTCATCCGCAAAGACGGCGAGTTCCGCCTG 3162
1326 er.AlaLysGlyGlnValAsnLeuSerAlaGlnAspGlySerValAlaGl 1342

3163 CATAATCCGGTCAAAGAACAAGAGCTTTCCGACAAACTCGGCAAGGCAGA 3212
1342 ySerIleAsnAlaAlaAsnValThrLeuAsnThrThrGlyThrLeuThrT 1359

3213 AGCCAAAAAACAGGCGGAAAAAGACAACGCGCAAAGCCTTG 3253
1359 hrValLysGlySerAsnIleAsnAlaThrSerGlyThrLeu 1372
```

seq_name: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-530-198-2

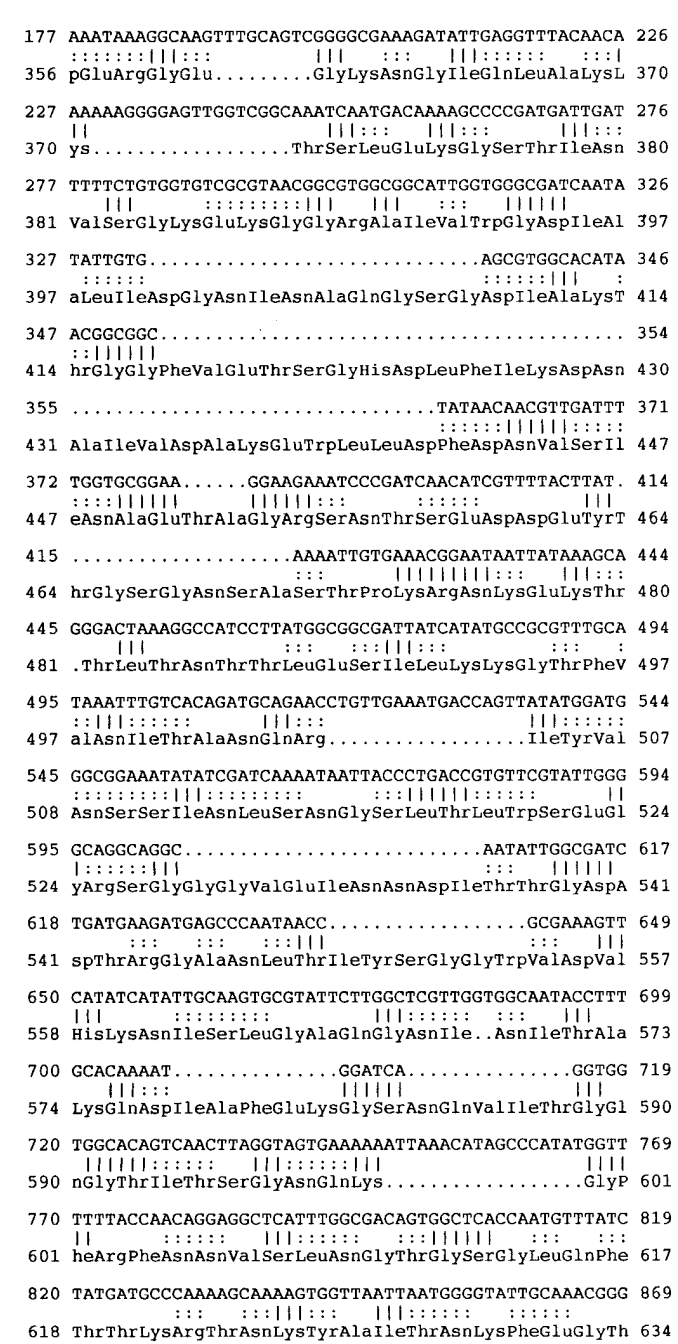
```
seq_documentation_block:
; Sequence 2, Application US/08530198
; Patent No. 5869065
;  GENERAL INFORMATION:
;    APPLICANT:  BARENKAMP, STEPHEN J
;    APPLICANT:  ST. GEME III, JOSEPH W
;    TITLE OF INVENTION:  HIGH MOLECULAR WEIGHT SURFACE PROTEINS
;    TITLE OF INVENTION:  OF NON-TYPEABLE HAEMOPHILUS
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Shoemaker and Mattare, Ltd
;      STREET:  2001 Jefferson Davis Hwy., 1203 Crystal Plaza
;      STREET:  Bldg. 1
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202-0286
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/530,198
;      FILING DATE:  13-DEC-1995
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BERKSTRESSER, JERRY W
;      REGISTRATION NUMBER:  22,651
;      REFERENCE/DOCKET NUMBER:  JWB-1186
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 415-0810
;      TELEFAX:  (703) 415-0813
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1536 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-530-198-2


alignment_scores:
            Quality:  272.50              Length:   1182
              Ratio:  0.452                 Gaps:     60
 Percent Similarity:  51.015   Percent Identity: 20.897

alignment_block:
 US-09-303-518D-649 x US-08-530-198-2   ..

 Align seg 1/1  to: US-08-530-198-2  from: 1  to: 1536
```

```
 127 GGACACACTTATTTCGGCATCAACTACCAATACTATCGCGACTTTGCCGA 176
 349 GlyGluThrTyrLeuGly.....GlyAs 356
```


177 AAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACA 226
356 pGluArgGlyGlu.....GlyLysAsnGlyIleGlnLeuAlaLysL 370
227 AAAAAGGGGAGTTGGTCGGCAAATCAATGACAAAAGCCCCGATGATTGAT 276
370 ys.....ThrSerLeuGluLysGlySerThrIleAsn 380
277 TTTTCTGTGGTGTCGCGTAACGGCGTGGCGGCATTGGTGGGCGATCAATA 326
381 ValSerGlyLysGluLysGlyGlyArgAlaIleValTrpGlyAspIleAl 397
327 TATTGTG.....AGCGTGGCACATA 346
397 aLeuIleAspGlyAsnIleAsnAlaGlnGlySerGlyAspIleAlaLysT 414
347 ACGGCGGC..... 354
414 hrGlyGlyPheValGluThrSerGlyHisAspLeuPheIleLysAspAsn 430
355 TATAACAACGTTGATTT 371
431 AlaIleValAspAlaLysGluTrpLeuLeuAspPheAspAsnValSerIl 447
372 TGGTGCGGAA.....GGAAGAAATCCCGATCAACATCGTTTTACTTAT. 414
447 eAsnAlaGluThrAlaGlyArgSerAsnThrSerGluAspAspGluTyrT 464
415 AAAATTGTGAAACGGAATAATTATAAAGCA 444
464 hrGlySerGlyAsnSerAlaSerThrProLysArgAsnLysGluLysThr 480
445 GGGACTAAAGGCCATCCTTATGGCGGCGATTATCATATGCCGCGTTTGCA 494
481 .ThrLeuThrAsnThrThrLeuGluSerIleLeuLysLysGlyThrPheV 497
495 TAAATTTGTCACAGATGCAGAACCTGTTGAAATGACCAGTTATATGGATG 544
497 alAsnIleThrAlaAsnGlnArg.....IleTyrVal 507
545 GGCGGAAATATATCGATCAAAATAATTACCCTGACCGTGTTCGTATTGGG 594
508 AsnSerSerIleAsnLeuSerAsnGlySerLeuThrLeuTrpSerGluGl 524
595 GCAGGCAGGC.....AATATTGGCGATC 617
524 yArgSerGlyGlyGlyValGluIleAsnAsnAspIleThrThrGlyAspA 541
618 TGATGAAGATGAGCCCAATAACC.....GCGAAAGTT 649
541 spThrArgGlyAlaAsnLeuThrIleTyrSerGlyGlyTrpValAspVal 557
650 CATATCATATTGCAAGTGCGTATTCTTGGCTCGTTGGTGGCAATACCTTT 699
558 HisLysAsnIleSerLeuGlyAlaGlnGlyAsnIle..AsnIleThrAla 573
700 GCACAAAAT.....GGATCA.....GGTGG 719
574 LysGlnAspIleAlaPheGluLysGlySerAsnGlnValIleThrGlyGl 590
720 TGGCACAGTCAACTTAGGTAGTGAAAAAATTAAACATAGCCCATATGGTT 769
590 nGlyThrIleThrSerGlyAsnGlnLys.....GlyP 601
770 TTTTACCAACAGGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATC 819
601 heArgPheAsnAsnValSerLeuAsnGlyThrGlySerGlyLeuGlnPhe 617
820 TATGATGCCCAAAAGCAAAAGTGGTTAATTAATGGGGTATTGCAAACGGG 869
618 ThrThrLysArgThrAsnLysTyrAlaIleThrAsnLysPheGluGlyTh 634
870 CAACCCCTATATAGGAAAAAGCAAT...GGCTTCCAGCTGGTTCGTAAAG 916
634 rLeuAsnIleSerGlyLysValAsnIleSerMetValLeuProLysAsnG 651
917 ATTGGTTCTATGATGAAATCTTTGCTGGAGATACCCAT.....TCAGTA 960
651 luSerGlyTyrAspLys...PheLysGlyArgThrTyrTrpAsnLeuThr 666
961 TTCTACGAACCACGTCAAAATGGGAAATACTCTTTTAACGACGATAATAA 1010
667 SerLeuAsnValSerGluSerGlyGluPheAsnLeuThrIleAspSerAr 683
1011 TGGCACA.....GGAAAAATCAATGCCAAACATGAACACAATTCTC 1051
683 gGlySerAspSerAlaGlyThrLeuThrGlnProTyrAsnLeuAsnGlyI 700
1052 TGCCT...AATAGATTAAAAACACGAACCGTTCAATTGTTTAATGTTTCT 1098
700 leSerPheAsnLysAspThrThrPheAsnValGluArg...AsnAlaArg 715
1099 TTATCCGAGACAGCAAGAGAACCTGTTTATCATGCTGCAGGTGGTGTCAA 1148
716 ValAsnPheAspIleLysAlaProIle.....GlyIleAs 727
1149 CAGTTATCGACCCAGACTGAATAATGGAGAAAATATTTCCTTTATTGACG 1198
727 nLysTyr.....SerSerLeuAsnTyrAlaSer. 736
1199 AAGGAAAAGGCGAATTGATACTTACCAGCAACATCAATCAAGGTGCTGGA 1248
737 PheAsnGlyAsnIleSerValSerGlyGly 746
1249 GGATTATATTTCCAAGGAGATTTTACGGTCTCGCCTGAAAATAACGAAAC 1298
747 Gly.....SerValAspPheThrLeuLeuAlaSerSerSerAsnVa 760
1299 TTGGCAAGGCGCGGGCGTT.....CATATCAGTG 1327
760 l...GlnThrProGlyValValIleAsnSerLysTyrPheAsnValSerT 776
1328 AAGACAGTACCGTTACTTGGAAAGTAAACGGC..... 1359
776 hrGlySerSerLeuArgPheLysThrSerGlySerThrLysThrGlyPhe 792
1360 ...GTGGCAAACGAC...CGCCTGTCCAAAATCGGCAAAGGCACGCTGCA 1403
793 SerIleGluLysAspLeuThrLeuAsnAlaThrGlyGlyAsnIleThrLe 809
1404 CGTTCAAGCCAAAGGGGAAAAC.....C 1426
809 uLeuGlnValGluGlyThrAspGlyMetIleGlyLysGlyIleValAlaL 826
1427 AAGGCTCGATCAGCGTGGGCGACGGTACAGTCATTTTGGATCAGCAGGCA 1476
826 ysLysAsnIleThrPheGluGlyGlyAsnIleThrPhe..... 838
1477 GACGATAAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCTTGGTCAGCGG 1526
839 GlySerArgLysAlaValThrGluIle.....GluGl 849
1527 CAGGGGTACGGTGCAACTGAATGCCGATAATCAGTTCAACCCCGACAAAC 1576
849 yAsnValThrIleAsnAsnAsnAlaAsn..... 858
1577 TCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAACGGGCATTCGCTT 1626
859 ValThrLeuIleGlySerAspPhe 866
1627 TCGTTCCACCGTATTCAAAATACCGATGAAGGGGCGATGATTGTCAACCA 1676
867 AspAsnHisGlnLysProLeuThrIleLysLysAspValIleIleAsnSe 883
1677 CAATCAA.....GACAAAGAATCCACCGTTACCATTACAGGCAATAAAG 1720

```
 883 rGlyAsnLeuThrAlaGlyGlyAsnIleValAsnIleAlaGlyAsnLeuT 900

1721 ATATTGCTACAACCGGCAAT..............................  1740
 900 hrValGluSerAsnAlaAsnPheLysAlaIleThrAsnPheThrPheAsn  916

1741 ..............AACAACAGCTTGGATAGCAAAAAAGAAATTGCCTA  1775
 917 ValGlyGlyLeuPheAspAsnLysGlyAsnSerAsnIleSerIleAlaLy  933

1776 CAACGGT...TGGTTTGGCGAGAAAGATACGACCAAAACGAACGGGCGGC  1822
 933 sGlyGlyAlaArgPheLysAspIleAspAsnSerLys.........AsnL  947

1823 TCAACCTTGTTTACCAGCCCGCCGCAGAAGACCGCACCCTGCTGCTTTCC  1872
 947 euSerIleThrThrAsnSerSerSerThrTyrArgThrIleIleSerGly  963

1873 GGCGGAACAAATTTAAACGGCAACATCACGCAAACAAACGGCAAACTGTT  1922
 964 AsnIleThrAsnLysAsnGlyAspLeuAsnIleThrAsn...........  976

1923 TTTCAGCGGCAGACCAACACCGCACGCCTACAATCATTTAAACGACCATT  1972
 977 ....GluGlySerAspThrGlu.........MetGlnIleGlyGlyAspV  989

1973 GGTCGCAAAAAGAGGGCATTCCTCGCGGGGAAATCGTGTGGGACAACGAC  2022
 989 alSerGlnLysGlu............GlyAsnLeuThrIleSerSerAsp  1001

2023 TGGATCAACCGCACATTTAAAGCGGAAAACTTCCAAATTAAAGGCGGACA  2072
1002 LysIleAsnIleThr.........LysGlnIleThrIleLysAlaGly..  1014

2073 GGCGGTGGTTTCCCGCAATGTTGCCAAAGTGAAAGGCGATTGGCATTTGA  2122
1015 ..................ValAspGlyGluAsnSerAspSerAspAlaT  1025

2123 GCAATCACGCCCAAGCAGTTTTTGGTGTCGCACCGCATCAAAGCCACACA  2172
1025 hrAsnAsnAlaAsnLeu.............................Thr  1031

2173 ATCTGTACACGTTCGGACTGGACGGGTCTGACAAATTGTGTCGAAAAAAC  2222
1032 IleLysThrLys.............................GluLeuLy  1038

2223 CATTACCGACGATAAAGTGATTGCTTCATTGACTAAGACCGACATCAGCG  2272
1038 sLeuThrGlnAspLeuAsnIleSerGlyPheAsnLysAlaGluIleThrA  1055

2273 GCAATGTCGATCTTGCCGATCACGCTCATTTAAATCTCACAGGGCTTGCC  2322
1055 laLys...AspGlySerAsp.........................Leu  1061

2323 ACACTCAACGGCAATCTTAGTGCAAATGGCGATACACGTTATACAGTCAG  2372
1062 ThrIleGlyAsnThrAsnSerAlaAspGlyThrAsnAlaLysLysValTh  1078

2373 CCACAACGCCACCCAAAACGGCAACCTTAGCCTCGTGGGCAATGCCCAAG  2422
1078 rPheAsnGlnValLysAspSerLysIleSerAlaAspGly..........  1091

2423 CAACATTTAATCAAGCCACATTAAACGGCAACACATCGGCTTCGGGCAAT  2472
1092 ........HisLysValThrLeuHisSerLysValGluThrSerGlySer  1105

2473 GCTTCATTTAATCTAAGCGACCACGCCGTACAAAACGGCAGTCTGACGCT  2522
1106 AsnAsn...AsnThrGluAspSerSerAspAsnAsnAlaGlyLeuThrIl  1121

2523 TTCCGGCAACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACGGTAATG  2572
```

```
1121 e......AspAlaLys...AsnValThr.........ValAsnAsnAsnI 1132

2573 TCTCCCTAGCCGATAAGGCAGTATTCCATTTTGAAAGCAGCCGCTTTACC 2622
1132 leThr...SerHisLysAlaVal.........SerIleSerAlaThrSer 1144

2623 GGACAAATCAGCGGCGGCAAGGATACGGCATTACACTTAAAA......GA 2666
1145 GlyGluIleThrThrLysThrGlyThrThrIleAsnAlaThrThrGlyAs 1161

2667 CAGCGAATGGACGCTGCCGTCAGGCACGGAATTAGGCAATTTAAACCTTG 2716
1161 nValGluIleThrAlaGlnThrGlySerIleLeuGlyGlyIleGluSerS 1178

2717 ACAACGCCACCATTACACTCAATTCCGCCTATCGCCACGATGCGGCAGGG 2766
1178 erSerGlySerValThrLeuThrAlaThrGluGlyAlaLeuAlaValSer 1194

2767 GCGCAAACCGGCAGTGCGACAGATGCGCCGCGCCGCCGTTCGCGCCGTTC 2816
1195 AsnIleSerGlyAsnThrValThrValThrAlaAsnSerGlyAlaLeuTh 1211

2817 GCGCCGTTCCCTATTATCCGTTACACCGCCAACTTCGGTAGAATCCCGTT 2866
1211 rThrLeuAlaGlySerThrIleLysGlyThrGluSerValThrThrSerS 1228

2867 TCAACACGCTGACGGTAAACGGCAAATTGAACGGTCAGGGAACA...... 2910
1228 erGlnSerGlyAspIleGlyGlyThrIleSerGly...GlyThrValGlu 1243

2911 TTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGACAAATTGAAGCT 2960
1244 ValLysAlaThrGluSerLeuThrThrGlnSerAsnSerLysIleLys.. 1259

2961 GGCGGAAAGTTCCGAAGGCACTTACACCTTGGCGGTC............. 2997
1260 .AlaThrThrGlyGluAlaAsnValThrSerAlaThrGlyThrIleGlyG 1276

2998 ...........AACAATACCGGCAACGAACCTGCAAGCCTCGAACAATTG 3036
1276 lyThrIleSerGlyAsnThrValAsnValThrAlaAsnAlaGlyAspLeu 1292

3037 ACGGTA.......................GTGGAAGGAAAAGACAACAA 3062
1293 ThrValGlyAsnGlyAlaGluIleAsnAlaThrGluGlyAlaAlaThrLe 1309

3063 ACCGCTGTCCGAAAACCTTAATTTCACCCTGCAAAACGAACACGTCGATG 3112
1309 uThrThrSerSerGlyLysLeuThrThrGluAlaSerSerHisIleThrS 1326

3113 CCGGCGCGTGGCGTTACCAACTCATCCGCAAAGACGGCGAGTTCCGCCTG 3162
1326 er.AlaLysGlyGlnValAsnLeuSerAlaGlnAspGlySerValAlaGl 1342

3163 CATAATCCGGTCAAAGAACAAGAGCTTTCCGACAAACTCGGCAAGGCAGA 3212
1342 ySerIleAsnAlaAlaAsnValThrLeuAsnThrThrGlyThrLeuThrT 1359

3213 AGCCAAAAAACAGGCGGAAAAAGACAACGCGCAAAGCCTTG 3253
1359 hrValLysGlySerAsnIleAsnAlaThrSerGlyThrLeu 1372
```

```
seq_name: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-469-880-2

seq_documentation_block:
; Sequence 2, Application US/08469880
; Patent No. 5876733
;  GENERAL INFORMATION:
;    APPLICANT:  Barenkamp, Stephen J.
;    TITLE OF INVENTION:  High Molecular Weight Surface Proteins
;    TITLE OF INVENTION:  of No. 5876733-Typeable Haemophilus
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
```

```
;       ADDRESSEE:  Shoemaker and Mattare, Ltd.
;       STREET:  2001 Jefferson Davis Hwy., 1203 Crystal Plaza
;       STREET:  Bldg. 1
;       CITY:  Arlington
;       STATE:  Virginia
;       COUNTRY:  U.S.A.
;       ZIP:  22202-0286
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/469,880
;       FILING DATE:  06-JUN-1995
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  GB 9205704.1
;       FILING DATE:  16-MAR-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US PCT/US93/02166
;       FILING DATE:  16-MAR-1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/302,832
;       FILING DATE:  16-SEP-1994
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Berkstresser, Jerry W
;       REGISTRATION NUMBER:  22,651
;       REFERENCE/DOCKET NUMBER:  1038-516 MIS:vg
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (703) 415-0810
;       TELEFAX:  (703) 415-0813
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1536 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
US-08-469-880-2


alignment_scores:
            Quality:  272.50              Length:   1182
              Ratio:  0.452                 Gaps:     60
 Percent Similarity:  51.015   Percent Identity: 20.897

alignment_block:
 US-09-303-518D-649 x US-08-469-880-2   ..

 Align seg 1/1  to: US-08-469-880-2  from: 1  to: 1536

     127 GGACACACTTATTTCGGCATCAACTACCAATACTATCGCGACTTTGCCGA 176
         |||:::|||||||:::|||                              :::::
     349 GlyGluThrTyrLeuGly.........................GlyAs 356

     177 AAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACA 226
         ::::::::|||:::         |||   :::   |||::::::   :::|
     356 pGluArgGlyGlu.........GlyLysAsnGlyIleGlnLeuAlaLysL 370

     227 AAAAAGGGGAGTTGGTCGGCAAATCAATGACAAAAGCCCCGATGATTGAT 276
         ||                        |||:::   |||:::      |||:::
     370 ys..................ThrSerLeuGluLysGlySerThrIleAsn 380

     277 TTTTCTGTGGTGTCGCGTAACGGCGTGGCGGCATTGGTGGGCGATCAATA 326
            |||      ::::::::::|||   |||   :::   ||||||
     381 ValSerGlyLysGluLysGlyGlyArgAlaIleValTrpGlyAspIleAl 397

     327 TATTGTG..............................AGCGTGGCACATA 346
          ::::::                             ::::::|||   :
     397 aLeuIleAspGlyAsnIleAsnAlaGlnGlySerGlyAspIleAlaLysT 414

     347 ACGGCGGC.......................................... 354

         ::|||||
414 hrGlyGlyPheValGluThrSerGlyHisAspLeuPheIleLysAspAsn 430

355 ..............................TATAACAACGTTGATTT 371
                                   ::::::|||||:::::
431 AlaIleValAspAlaLysGluTrpLeuLeuAspPheAspAsnValSerIl 447

372 TGGTGCGGAA......GGAAGAAATCCCGATCAACATCGTTTTACTTAT. 414
    ::::||||||      |||||||:::      ::::::         |||
447 eAsnAlaGluThrAlaGlyArgSerAsnThrSerGluAspAspGluTyrT 464

415 ...................AAAATTGTGAAACGGAATAATTATAAAGCA 444
                        :::      |||||||||:::   |||:::
464 hrGlySerGlyAsnSerAlaSerThrProLysArgAsnLysGluLysThr 480

445 GGGACTAAAGGCCATCCTTATGGCGGCGATTATCATATGCCGCGTTTGCA 494
      |||          :::   :::|||:::         :::   :
481 .ThrLeuThrAsnThrThrLeuGluSerIleLeuLysLysGlyThrPheV 497

495 TAAATTTGTCACAGATGCAGAACCTGTTGAAATGACCAGTTATATGGATG 544
    ::|||::::::      |||:::                   |||::::::
497 alAsnIleThrAlaAsnGlnArg.................IleTyrVal 507

545 GGCGGAAATATATCGATCAAAATAATTACCCTGACCGTGTTCGTATTGGG 594
    :::::::::|||:::::::::      :::||||||::::::      ||
508 AsnSerSerIleAsnLeuSerAsnGlySerLeuThrLeuTrpSerGluGl 524

595 GCAGGCAGGC.........................AATATTGGCGATC 617
    |::::::|||                         :::   ||||||
524 yArgSerGlyGlyGlyValGluIleAsnAsnAspIleThrThrGlyAspA 541

618 TGATGAAGATGAGCCCAATAACC..................GCGAAAGTT 649
        :::   :::   :::|||                   :::   |||
541 spThrArgGlyAlaAsnLeuThrIleTyrSerGlyGlyTrpValAspVal 557

650 CATATCATATTGCAAGTGCGTATTCTTGGCTCGTTGGTGGCAATACCTTT 699
    |||      :::::::::         |||::::::  :::   |||
558 HisLysAsnIleSerLeuGlyAlaGlnGlyAsnIle..AsnIleThrAla 573

700 GCACAAAAT..............GGATCA..............GGTGG 719
       |||:::               ||||||               |||
574 LysGlnAspIleAlaPheGluLysGlySerAsnGlnValIleThrGlyGl 590

720 TGGCACAGTCAACTTAGGTAGTGAAAAAATTAAACATAGCCCATATGGTT 769
     ||||||::::::   |||:::::|||                    ||||
590 nGlyThrIleThrSerGlyAsnGlnLys.................GlyP 601

770 TTTTACCAACAGGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATC 819
    ||      ::::::   |||::::::   :::||||||   :::   :::
601 heArgPheAsnAsnValSerLeuAsnGlyThrGlySerGlyLeuGlnPhe 617

820 TATGATGCCCAAAAGCAAAAGTGGTTAATTAATGGGGTATTGCAAACGGG 869
              :::   :::|||:::   |||::::::   ::::::
618 ThrThrLysArgThrAsnLysTyrAlaIleThrAsnLysPheGluGlyTh 634

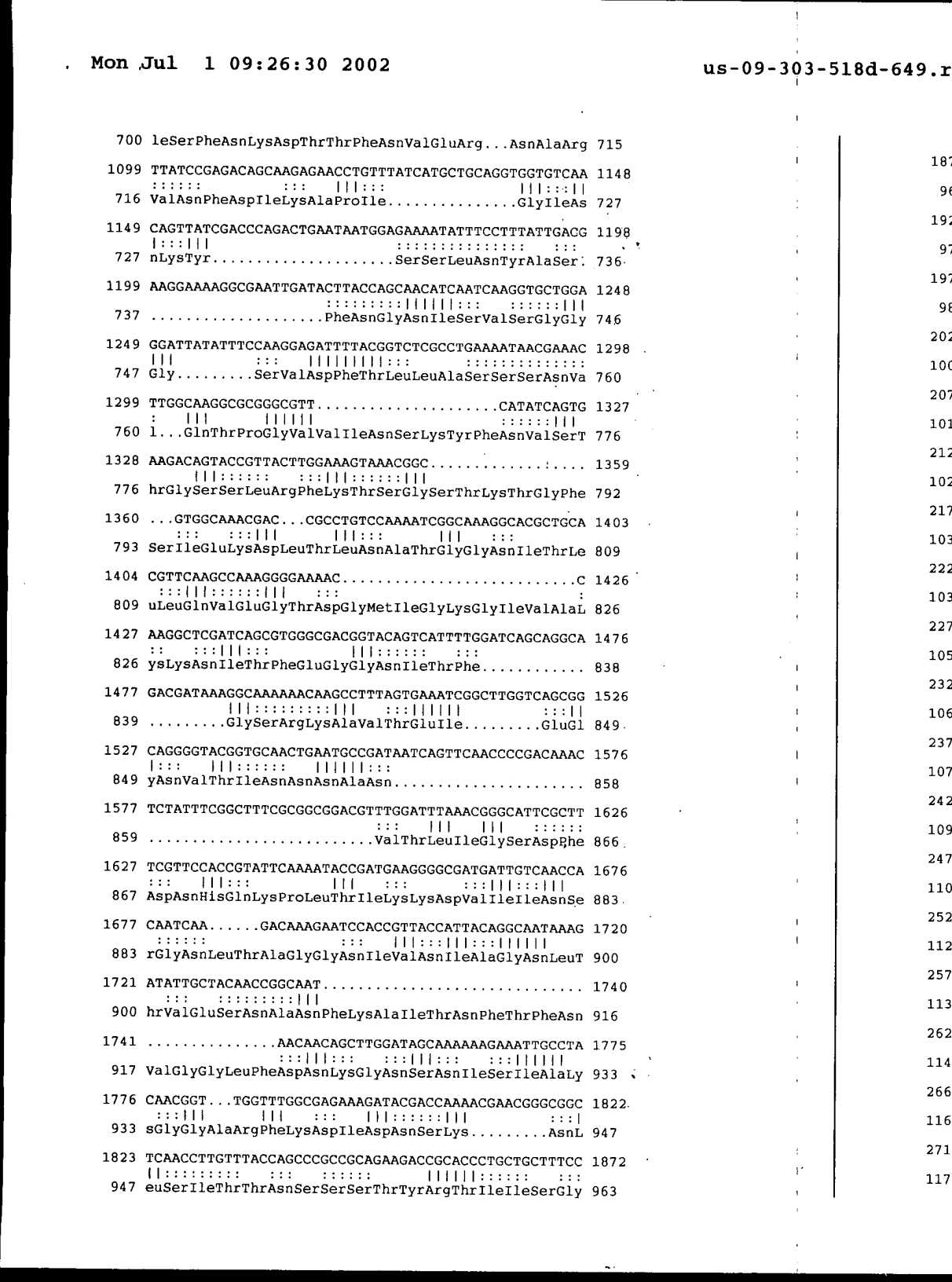
870 CAACCCCTATATAGGAAAAAGCAAT...GGCTTCCAGCTGGTTCGTAAAG 916
                  ||||||   |||   ::::::   |||   :::::::
634 rLeuAsnIleSerGlyLysValAsnIleSerMetValLeuProLysAsnG 651

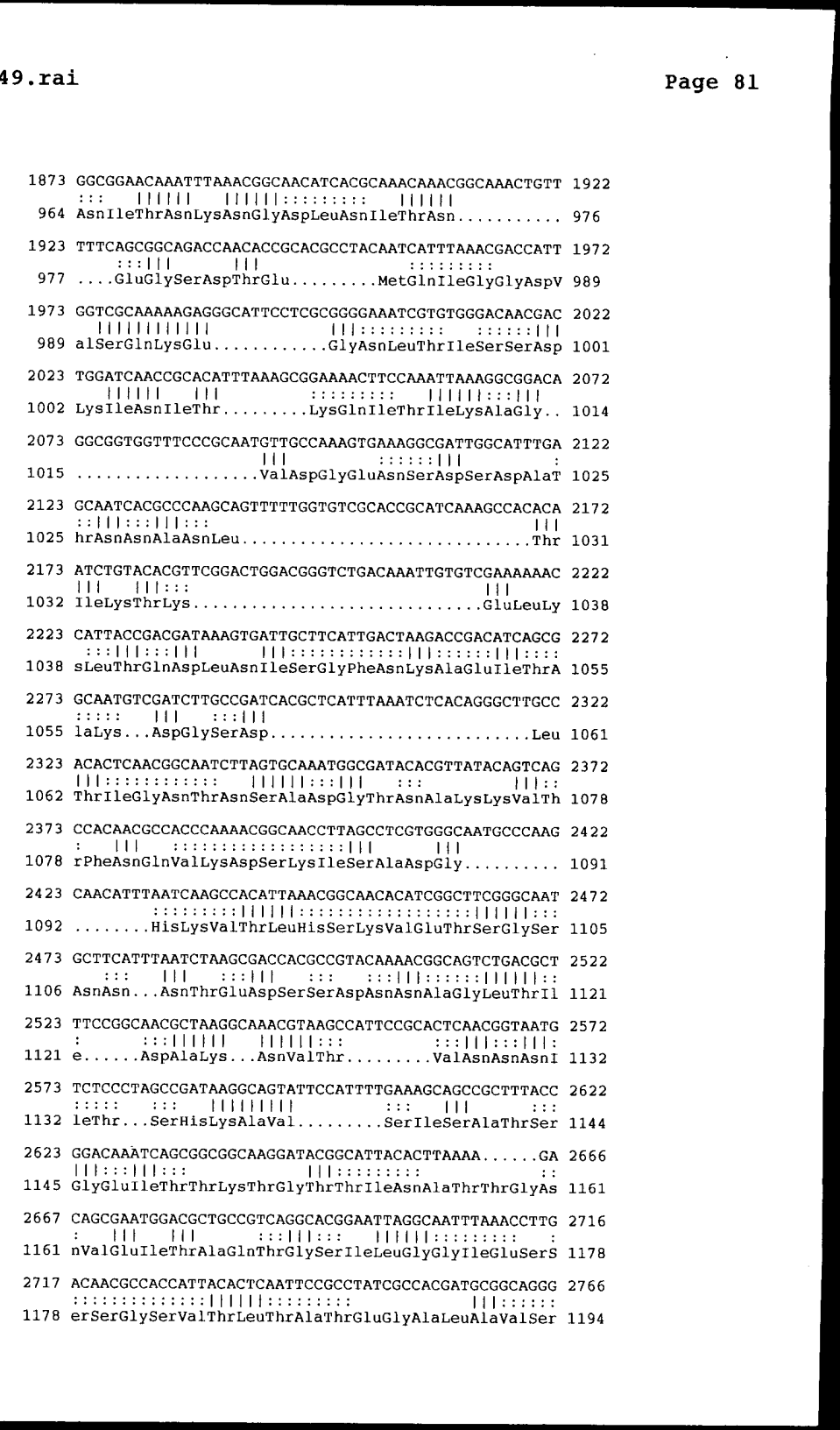
917 ATTGGTTCTATGATGAAATCTTTGCTGGAGATACCCAT......TCAGTA 960
    ::      |||||||:::   |||   |||   |||:::         :::
651 luSerGlyTyrAspLys...PheLysGlyArgThrTyrTrpAsnLeuThr 666

961 TTCTACGAACCACGTCAAAATGGGAAATACTCTTTTAACGACGATAATAA 1010
          :::      ::::::|||:::::::::::::::::   |||:::::
667 SerLeuAsnValSerGluSerGlyGluPheAsnLeuThrIleAspSerAr 683

1011 TGGCACA.........GGAAAAATCAATGCCAAACATGAACACAATTCTC 1051
     :|||:::         |||   ::::::      ::::::   |||::::
683  gGlySerAspSerAlaGlyThrLeuThrGlnProTyrAsnLeuAsnGlyI 700

1052 TGCCT...AATAGATTAAAAACACGAACCGTTCAATTGTTTAATGTTTCT 1098
     ::      |||:::      |||   :::|||:::      |||:::
```


700 leSerPheAsnLysAspThrThrPheAsnValGluArg...AsnAlaArg 715
1099 TTATCCGAGACAGCAAGAGAACCTGTTTATCATGCTGCAGGTGGTGTCAA 1148
716 ValAsnPheAspIleLysAlaProIle.....GlyIleAs 727
1149 CAGTTATCGACCCAGACTGAATAATGGAGAAAATATTTCCTTTATTGACG 1198
727 nLysTyr.....SerSerLeuAsnTyrAlaSer 736
1199 AAGGAAAAGGCGAATTGATACTTACCAGCAACATCAATCAAGGTGCTGGA 1248
737 PheAsnGlyAsnIleSerValSerGlyGly 746
1249 GGATTATATTTCCAAGGAGATTTTACGGTCTCGCCTGAAAATAACGAAAC 1298
747 Gly.....SerValAspPheThrLeuLeuAlaSerSerSerAsnVa 760
1299 TTGGCAAGGCGCGGGCGTT.....CATATCAGTG 1327
760 l...GlnThrProGlyValValIleAsnSerLysTyrPheAsnValSerT 776
1328 AAGACAGTACCGTTACTTGGAAAGTAAACGGC..... 1359
776 hrGlySerSerLeuArgPheLysThrSerGlySerThrLysThrGlyPhe 792
1360 ...GTGGCAAACGAC...CGCCTGTCCAAAATCGGCAAAGGCACGCTGCA 1403
793 SerIleGluLysAspLeuThrLeuAsnAlaThrGlyGlyAsnIleThrLe 809
1404 CGTTCAAGCCAAAGGGGAAAAC.....C 1426
809 uLeuGlnValGluGlyThrAspGlyMetIleGlyLysGlyIleValAlaL 826
1427 AAGGCTCGATCAGCGTGGGCGACGGTACAGTCATTTTGGATCAGCAGGCA 1476
826 ysLysAsnIleThrPheGluGlyGlyAsnIleThrPhe..... 838
1477 GACGATAAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCTTGGTCAGCGG 1526
839 GlySerArgLysAlaValThrGluIle.....GluGl 849
1527 CAGGGGTACGGTGCAACTGAATGCCGATAATCAGTTCAACCCCGACAAAC 1576
849 yAsnValThrIleAsnAsnAsnAlaAsn..... 858
1577 TCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAACGGGCATTCGCTT 1626
859 ValThrLeuIleGlySerAspPhe 866
1627 TCGTTCCACCGTATTCAAAATACCGATGAAGGGGCGATGATTGTCAACCA 1676
867 AspAsnHisGlnLysProLeuThrIleLysLysAspValIleIleAsnSe 883
1677 CAATCAA.....GACAAAGAATCCACCGTTACCATTACAGGCAATAAAG 1720
883 rGlyAsnLeuThrAlaGlyGlyAsnIleValAsnIleAlaGlyAsnLeuT 900
1721 ATATTGCTACAACCGGCAAT..... 1740
900 hrValGluSerAsnAlaAsnPheLysAlaIleThrAsnPheThrPheAsn 916
1741 AACAACAGCTTGGATAGCAAAAAAGAAATTGCCTA 1775
917 ValGlyGlyLeuPheAspAsnLysGlyAsnSerAsnIleSerIleAlaLy 933
1776 CAACGGT...TGGTTTGGCGAGAAAGATACGACCAAAACGAACGGGCGGC 1822
933 sGlyGlyAlaArgPheLysAspIleAspAsnSerLys.....AsnL 947
1823 TCAACCTTGTTTACCAGCCCGCCGCAGAAGACCGCACCCTGCTGCTTTCC 1872
947 euSerIleThrThrAsnSerSerSerThrTyrArgThrIleIleSerGly 963


1873 GGCGGAACAAATTTAAACGGCAACATCACGCAAACAAACGGCAAACTGTT 1922
964 AsnIleThrAsnLysAsnGlyAspLeuAsnIleThrAsn..... 976
1923 TTTCAGCGGCAGACCAACACCGCACGCCTACAATCATTTAAACGACCATT 1972
977 GluGlySerAspThrGlu.....MetGlnIleGlyGlyAspV 989
1973 GGTCGCAAAAAGAGGGCATTCCTCGCGGGGAAATCGTGTGGGACAACGAC 2022
989 alSerGlnLysGlu.....GlyAsnLeuThrIleSerSerAsp 1001
2023 TGGATCAACCGCACATTTAAAGCGGAAAACTTCCAAATTAAAGGCGGACA 2072
1002 LysIleAsnIleThr.....LysGlnIleThrIleLysAlaGly.. 1014
2073 GGCGGTGGTTTCCCGCAATGTTGCCAAAGTGAAAGGCGATTGGCATTTGA 2122
1015 ValAspGlyGluAsnSerAspSerAspAlaT 1025
2123 GCAATCACGCCCAAGCAGTTTTTGGTGTCGCACCGCATCAAAGCCACACA 2172
1025 hrAsnAsnAlaAsnLeu.....Thr 1031
2173 ATCTGTACACGTTCGGACTGGACGGGTCTGACAAATTGTGTCGAAAAAAC 2222
1032 IleLysThrLys.....GluLeuLy 1038
2223 CATTACCGACGATAAAGTGATTGCTTCATTGACTAAGACCGACATCAGCG 2272
1038 sLeuThrGlnAspLeuAsnIleSerGlyPheAsnLysAlaGluIleThrA 1055
2273 GCAATGTCGATCTTGCCGATCACGCTCATTTAAATCTCACAGGGCTTGCC 2322
1055 laLys...AspGlySerAsp.....Leu 1061
2323 ACACTCAACGGCAATCTTAGTGCAAATGGCGATACACGTTATACAGTCAG 2372
1062 ThrIleGlyAsnThrAsnSerAlaAspGlyThrAsnAlaLysLysValTh 1078
2373 CCACAACGCCACCCAAAACGGCAACCTTAGCCTCGTGGGCAATGCCCAAG 2422
1078 rPheAsnGlnValLysAspSerLysIleSerAlaAspGly..... 1091
2423 CAACATTTAATCAAGCCACATTAAACGGCAACACATCGGCTTCGGGCAAT 2472
1092 HisLysValThrLeuHisSerLysValGluThrSerGlySer 1105
2473 GCTTCATTTAATCTAAGCGACCACGCCGTACAAAACGGCAGTCTGACGCT 2522
1106 AsnAsn...AsnThrGluAspSerSerAspAsnAsnAlaGlyLeuThrIl 1121
2523 TTCCGGCAACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACGGTAATG 2572
1121 e.....AspAlaLys...AsnValThr.....ValAsnAsnAsnI 1132
2573 TCTCCCTAGCCGATAAGGCAGTATTCCATTTTGAAAGCAGCCGCTTTACC 2622
1132 leThr...SerHisLysAlaVal.....SerIleSerAlaThrSer 1144
2623 GGACAAATCAGCGGCGGCAAGGATACGGCATTACACTTAAAA.....GA 2666
1145 GlyGluIleThrThrLysThrGlyThrThrIleAsnAlaThrThrGlyAs 1161
2667 CAGCGAATGGACGCTGCCGTCAGGCACGGAATTAGGCAATTTAAACCTTG 2716
1161 nValGluIleThrAlaGlnThrGlySerIleLeuGlyGlyIleGluSerS 1178
2717 ACAACGCCACCATTACACTCAATTCCGCCTATCGCCACGATGCGGCAGGG 2766
1178 erSerGlySerValThrLeuThrAlaThrGluGlyAlaLeuAlaValSer 1194

```
    2767 GCGCAAACCGGCAGTGCGACAGATGCGCCGCGCCGCCGTTCGCGCCGTTC 2816
    1195 AsnIleSerGlyAsnThrValThrValThrAlaAsnSerGlyAlaLeuTh 1211

    2817 GCGCCGTTCCCTATTATCCGTTACACCGCCAACTTCGGTAGAATCCCGTT 2866
    1211 rThrLeuAlaGlySerThrIleLysGlyThrGluSerValThrThrSerS 1228

    2867 TCAACACGCTGACGGTAAACGGCAAATTGAACGGTCAGGGAACA..... 2910
    1228 erGlnSerGlyAspIleGlyGlyThrIleSerGly...GlyThrValGlu 1243

    2911 TTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGACAAATTGAAGCT 2960
    1244 ValLysAlaThrGluSerLeuThrThrGlnSerAsnSerLysIleLys.. 1259

    2961 GGCGGAAAGTTCCGAAGGCACTTACACCTTGGCGGTC..... 2997
    1260 .AlaThrThrGlyGluAlaAsnValThrSerAlaThrGlyThrIleGlyG 1276

    2998 ..........AACAATACCGGCAACGAACCTGCAAGCCTCGAACAATTG 3036
    1276 lyThrIleSerGlyAsnThrValAsnValThrAlaAsnAlaGlyAspLeu 1292

    3037 ACGGTA.....................GTGGAAGGAAAAGACAACAA 3062
    1293 ThrValGlyAsnGlyAlaGluIleAsnAlaThrGluGlyAlaAlaThrLe 1309

    3063 ACCGCTGTCCGAAAACCTTAATTTCACCCTGCAAAACGAACACGTCGATG 3112
    1309 uThrThrSerSerGlyLysLeuThrThrGluAlaSerSerHisIleThrS 1326

    3113 CCGGCGCGTGGCGTTACCAACTCATCCGCAAAGACGGCGAGTTCCGCCTG 3162
    1326 er.AlaLysGlyGlnValAsnLeuSerAlaGlnAspGlySerValAlaGl 1342

    3163 CATAATCCGGTCAAAGAACAAGAGCTTTCCGACAAACTCGGCAAGGCAGA 3212
    1342 ySerIleAsnAlaAlaAsnValThrLeuAsnThrThrGlyThrLeuThrT 1359

    3213 AGCCAAAAAACAGGCGGAAAAAGACAACGCGCAAAGCCTTG 3253
    1359 hrValLysGlySerAsnIleAsnAlaThrSerGlyThrLeu 1372

seq_name: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-728-470-2

seq_documentation_block:
; Sequence 2, Application US/08728470
; Patent No. 5928651
;  GENERAL INFORMATION:
;    APPLICANT:  Barenkamp, Stephen J
;    TITLE OF INVENTION:  High Molecular Weight Surface Proteins
;    TITLE OF INVENTION:  of No. 5928651-Typeable Haemophilus
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Shoemaker and Mattare, Ltd.
;      STREET:  2001 Jefferson Davis Hwy., 1203 Crystal Plaza
;      STREET:  Bldg. 1
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202-0286
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/728,470
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/302,832
;      FILING DATE:  16-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US PCT/US93/02166
;      FILING DATE:  16-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9205704.1
;      FILING DATE:  16-MAR-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Berkstresser, Jerry W
;      REGISTRATION NUMBER:  22,651
;      REFERENCE/DOCKET NUMBER:  1038-633
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 415-0810
;      TELEFAX:  (703) 415-0813
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1536 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:
US-08-728-470-2


alignment_scores:
           Quality:  272.50            Length:   1182
             Ratio:  0.452               Gaps:     60
 Percent Similarity:  51.015   Percent Identity: 20.897

alignment_block:
 US-09-303-518D-649 x US-08-728-470-2   ..

 Align seq 1/1  to: US-08-728-470-2  from: 1  to: 1536

     127 GGACACACTTATTTCGGCATCAACTACCAATACTATCGCGACTTTGCCGA 176
     349 GlyGluThrTyrLeuGly.....GlyAs 356

     177 AAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACA 226
     356 pGluArgGlyGlu.....GlyLysAsnGlyIleGlnLeuAlaLysL 370

     227 AAAAAGGGGAGTTGGTCGGCAAATCAATGACAAAAGCCCCGATGATTGAT 276
     370 ys.....ThrSerLeuGluLysGlySerThrIleAsn 380

     277 TTTTCTGTGGTGTCGCGTAACGGCGTGGCGGCATTGGTGGGCGATCAATA 326
     381 ValSerGlyLysGluLysGlyGlyArgAlaIleValTrpGlyAspIleAl 397

     327 TATTGTG.....AGCGTGGCACATA 346
     397 aLeuIleAspGlyAsnIleAsnAlaGlnGlySerGlyAspIleAlaLysT 414

     347 ACGGCGGC..... 354
     414 hrGlyGlyPheValGluThrSerGlyHisAspLeuPheIleLysAspAsn 430

     355 .....TATAACAACGTTGATTT 371
     431 AlaIleValAspAlaLysGluTrpLeuLeuAspPheAspAsnValSerIl 447

     372 TGGTGCGGAA......GGAAGAAATCCCGATCAACATCGTTTTACTTAT. 414
     447 eAsnAlaGluThrAlaGlyArgSerAsnThrSerGluAspAspGluTyrT 464

     415 .....AAAATTGTGAAACGGAATAATTATAAAGCA 444
     464 hrGlySerGlyAsnSerAlaSerThrProLysArgAsnLysGluLysThr 480

     445 GGGACTAAAGGCCATCCTTATGGCGGCGATTATCATATGCCGCGTTTGCA 494
     481 .ThrLeuThrAsnThrThrLeuGluSerIleLeuLysLysGlyThrPheV 497
```



```
                        |||         ::::::|||         :
1015 ..................ValAspGlyGluAsnSerAspSerAspAlaT 1025

2123 GCAATCACGCCCAAGCAGTTTTTGGTGTCGCACCGCATCAAAGCCACACA 2172
     ::|||:::|||:::                                 |||
1025 hrAsnAsnAlaAsnLeu............................Thr 1031

2173 ATCTGTACACGTTCGGACTGGACGGGTCTGACAAATTGTGTCGAAAAAAC 2222
     |||   |||:::                                 |||
1032 IleLysThrLys............................GluLeuLy 1038

2223 CATTACCGACGATAAAGTGATTGCTTCATTGACTAAGACCGACATCAGCG 2272
      :::|||:::|||      |||:::::::::::::|||:::::::|||::::
1038 sLeuThrGlnAspLeuAsnIleSerGlyPheAsnLysAlaGluIleThrA 1055

2273 GCAATGTCGATCTTGCCGATCACGCTCATTTAAATCTCACAGGGCTTGCC 2322
     :::::   |||   :::|||
1055 laLys...AspGlySerAsp.........................Leu 1061

2323 ACACTCAACGGCAATCTTAGTGCAAATGGCGATACACGTTATACAGTCAG 2372
     |||:::::::::::::   ||||||:::|||   :::         |||::
1062 ThrIleGlyAsnThrAsnSerAlaAspGlyThrAsnAlaLysLysValTh 1078

2373 CCACAACGCCACCCAAAACGGCAACCTTAGCCTCGTGGGCAATGCCCAAG 2422
     :   |||   :::::::::::::::::|||      |||
1078 rPheAsnGlnValLysAspSerLysIleSerAlaAspGly.......... 1091

2423 CAACATTTAATCAAGCCACATTAAACGGCAACACATCGGCTTCGGGCAAT 2472
              :::::::::||||||:::::::::::::::::||||||:::
1092 ........HisLysValThrLeuHisSerLysValGluThrSerGlySer 1105

2473 GCTTCATTTAATCTAAGCGACCACGCCGTACAAAACGGCAGTCTGACGCT 2522
        :::   |||   :::|||   :::   :::|||:::::||||||::
1106 AsnAsn...AsnThrGluAspSerSerAspAsnAsnAlaGlyLeuThrIl 1121

2523 TTCCGGCAACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACGGTAATG 2572
     :      :::||||||   ||||||:::         :::|||:::|||:
1121 e......AspAlaLys...AsnValThr.........ValAsnAsnAsnI 1132

2573 TCTCCCTAGCCGATAAGGCAGTATTCCATTTTGAAAGCAGCCGCTTTACC 2622
     :::::   :::   |||||||||         :::   |||      :::
1132 leThr...SerHisLysAlaVal.........SerIleSerAlaThrSer 1144

2623 GGACAAATCAGCGGCGGCAAGGATACGGCATTACACTTAAAA......GA 2666
     |||:::|||:::           |||::::::::::          ::
1145 GlyGluIleThrThrLysThrGlyThrThrIleAsnAlaThrThrGlyAs 1161

2667 CAGCGAATGGACGCTGCCGTCAGGCACGGAATTAGGCAATTTAAACCTTG 2716
     :   |||   |||      :::|||:::   ||||||:::::::::   :
1161 nValGluIleThrAlaGlnThrGlySerIleLeuGlyGlyIleGluSerS 1178

2717 ACAACGCCACCATTACACTCAATTCCGCCTATCGCCACGATGCGGCAGGG 2766
     ::::::::::::::|||||:::::::::          |||::::::
1178 erSerGlySerValThrLeuThrAlaThrGluGlyAlaLeuAlaValSer 1194

2767 GCGCAAACCGGCAGTGCGACAGATGCGCCGCGCCGCCGTTCGCGCCGTTC 2816
           :::|||:::::::::   :::      :::   :::      ::
1195 AsnIleSerGlyAsnThrValThrValThrAlaAsnSerGlyAlaLeuTh 1211

2817 GCGCCGTTCCCTATTATCCGTTACACCGCCAACTTCGGTAGAATCCCGTT 2866
     :      :::      ::::::            ||||||   :::
1211 rThrLeuAlaGlySerThrIleLysGlyThrGluSerValThrThrSerS 1228

2867 TCAACACGCTGACGGTAAACGGCAAATTGAACGGTCAGGGAACA...... 2910
       ::::::      :::::::|||   ::::::|||   ||||||
1228 erGlnSerGlyAspIleGlyGlyThrIleSerGly...GlyThrValGlu 1243

2911 TTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGACAAATTGAAGCT 2960
        :::      ::::::|||            :::::::|||:::|||
1244 ValLysAlaThrGluSerLeuThrThrGlnSerAsnSerLysIleLys.. 1259

2961 GGCGGAAAGTTCCGAAGGCACTTACACCTTGGCGGTC.............. 2997
      |||   ::::::|||::::::   |||   |||:::
```

```
     1260 .AlaThrThrGlyGluAlaAsnValThrSerAlaThrGlyThrIleGlyG 1276

     2998 ..........AACAATACCGGCAACGAACCTGCAAGCCTCGAACAATTG 3036
                       :::||||||   |||      |||:::      :::|||
     1276 lyThrIleSerGlyAsnThrValAsnValThrAlaAsnAlaGlyAspLeu 1292

     3037 ACGGTA.....................GTGGAAGGAAAAGACAACAA 3062
          ||||||                    :::||||||      :::
     1293 ThrValGlyAsnGlyAlaGluIleAsnAlaThrGluGlyAlaAlaThrLe 1309

     3063 ACCGCTGTCCGAAAACCTTAATTTCACCCTGCAAAACGAACACGTCGATG 3112
                |||::::::         |||      ::::::|||:::   :
     1309 uThrThrSerSerGlyLysLeuThrThrGluAlaSerSerHisIleThrS 1326

     3113 CCGGCGCGTGGCGTTACCAACTCATCCGCAAAGACGGCGAGTTCCGCCTG 3162
          :: |||:::|||   :::|||   ||||||:::   :::|||   |||
     1326 er.AlaLysGlyGlnValAsnLeuSerAlaGlnAspGlySerValAlaGl 1342

     3163 CATAATCCGGTCAAAGAACAAGAGCTTTCCGACAAACTCGGCAAGGCAGA 3212
               |||::::::   |||   ::::::   |||:::::::::
     1342 ySerIleAsnAlaAlaAsnValThrLeuAsnThrThrGlyThrLeuThrT 1359

     3213 AGCCAAAAAACAGGCGGAAAAAGACAACGCGCAAAGCCTTG 3253
               |||:::   :::   ::::::|||      :::|||
     1359 hrValLysGlySerAsnIleAsnAlaThrSerGlyThrLeu 1372
```

seq_name: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-617-697-2

```
seq_documentation_block:
; Sequence 2, Application US/08617697
; Patent No. 5977336
;  GENERAL INFORMATION:
;    APPLICANT:  Barenkamp, Stephen J
;    TITLE OF INVENTION:  High Molecular Weight Surface Proteins
;    TITLE OF INVENTION:  of No. 5977336-Typeable Haemophilus
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Shoemaker and Mattare, Ltd.
;      STREET:  2001 Jefferson Davis Hwy., 1203 Crystal Plaza
;      STREET:  Bldg. 1
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202-0286
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/617,697
;      FILING DATE:  01-APR-1996
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/302,832
;      FILING DATE:  05-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US PCT/US93/02166
;      FILING DATE:  16-MAR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Berkstresser, Jerry W
;      REGISTRATION NUMBER:  22,651
;      REFERENCE/DOCKET NUMBER:  1038-557
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 415-0810
;      TELEFAX:  (703) 415-0813
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1536 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
```

US-08-617-697-2

alignment_scores:
Quality: 272.50 Length: 1182
Ratio: 0.452 Gaps: 60
Percent Similarity: 51.015 Percent Identity: 20.897

alignment_block:
US-09-303-518D-649 x US-08-617-697-2 ..

Align seg 1/1 to: US-08-617-697-2 from: 1 to: 1536

```
 127 GGACACACTTATTTCGGCATCAACTACCAATACTATCGCGACTTTGCCGA 176
     |||:::|||||||:::|||                              :::::
 349 GlyGluThrTyrLeuGly.........................GlyAs 356

 177 AAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACA 226
     :::::::|||:::          |||   :::   |||::::::   :::|
 356 pGluArgGlyGlu.........GlyLysAsnGlyIleGlnLeuAlaLysL 370

 227 AAAAAGGGGAGTTGGTCGGCAAATCAATGACAAAAGCCCCGATGATTGAT 276
     ||                     |||:::   |||:::      |||:::
 370 ys..................ThrSerLeuGluLysGlySerThrIleAsn 380

 277 TTTTCTGTGGTGTCGCGTAACGGCGTGGCGGCATTGGTGGGCGATCAATA 326
        |||      ::::::::::|||   |||   :::   ||||||
 381 ValSerGlyLysGluLysGlyGlyArgAlaIleValTrpGlyAspIleAl 397

 327 TATTGTG..............................AGCGTGGCACATA 346
      ::::::                             ::::::|||   :
 397 aLeuIleAspGlyAsnIleAsnAlaGlnGlySerGlyAspIleAlaLysT 414

 347 ACGGCGGC.......................................... 354
     ::||||||
 414 hrGlyGlyPheValGluThrSerGlyHisAspLeuPheIleLysAspAsn 430

 355 ................................TATAACAACGTTGATTT 371
                                    ::::::||||||:::::
 431 AlaIleValAspAlaLysGluTrpLeuLeuAspPheAspAsnValSerIl 447

 372 TGGTGCGGAA......GGAAGAAATCCCGATCAACATCGTTTTACTTAT. 414
     ::::||||||      ||||||:::      ::::::          |||
 447 eAsnAlaGluThrAlaGlyArgSerAsnThrSerGluAspAspGluTyrT 464

 415 ...................AAAATTGTGAAACGGAATAATTATAAAGCA 444
                         :::    |||||||||:::   |||:::
 464 hrGlySerGlyAsnSerAlaSerThrProLysArgAsnLysGluLysThr 480

 445 GGGACTAAAGGCCATCCTTATGGCGGCGATTATCATATGCCGCGTTTGCA 494
        |||          :::   :::|||:::         :::   :
 481 .ThrLeuThrAsnThrThrLeuGluSerIleLeuLysLysGlyThrPheV 497

 495 TAAATTTGTCACAGATGCAGAACCTGTTGAAATGACCAGTTATATGGATG 544
     ::|||::::::      |||:::                  |||::::::
 497 alAsnIleThrAlaAsnGlnArg.................IleTyrVal 507

 545 GGCGGAAATATATCGATCAAAATAATTACCCTGACCGTGTTCGTATTGGG 594
     :::::::::||::::::::::      :::||||||::::::      ||
 508 AsnSerSerIleAsnLeuSerAsnGlySerLeuThrLeuTrpSerGluGl 524

 595 GCAGGCAGGC..........................AATATTGGCGATC 617
     |:::::|||                          :::   ||||||
 524 yArgSerGlyGlyGlyValGluIleAsnAsnAspIleThrThrGlyAspA 541

 618 TGATGAAGATGAGCCCAATAACC.................GCGAAAGTT 649
         :::   :::   :::|||                  :::   |||
 541 spThrArgGlyAlaAsnLeuThrIleTyrSerGlyGlyTrpValAspVal 557

 650 CATATCATATTGCAAGTGCGTATTCTTGGCTCGTTGGTGGCAATACCTTT 699
     |||      :::::::::         |||::::::  :::   |||
 558 HisLysAsnIleSerLeuGlyAlaGlnGlyAsnIle..AsnIleThrAla 573

 700 GCACAAAAT..............GGATCA..............GGTGG 719
        |||:::               ||||||               |||
 574 LysGlnAspIleAlaPheGluLysGlySerAsnGlnValIleThrGlyGl 590

 720 TGGCACAGTCAACTTAGGTAGTGAAAAAATTAAACATAGCCCATATGGTT 769
      ||||||::::::   |||:::::|||                    ||||
 590 nGlyThrIleThrSerGlyAsnGlnLys.................GlyP 601

 770 TTTTACCAACAGGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATC 819
     ||      ::::::   |||::::::   :::||||||   :::   :::
 601 heArgPheAsnAsnValSerLeuAsnGlyThrGlySerGlyLeuGlnPhe 617

 820 TATGATGCCCAAAAGCAAAAGTGGTTAATTAATGGGGTATTGCAAACGGG 869
               :::   :::|||:::   |||::::::   ::::::
 618 ThrThrLysArgThrAsnLysTyrAlaIleThrAsnLysPheGluGlyTh 634

 870 CAACCCCTATATAGGAAAAAGCAAT...GGCTTCCAGCTGGTTCGTAAAG 916
                  ||||||   |||   ::::::   |||   :::::::
 634 rLeuAsnIleSerGlyLysValAsnIleSerMetValLeuProLysAsnG 651

 917 ATTGGTTCTATGATGAAATCTTTGCTGGAGATACCCAT......TCAGTA 960
     ::      ||||||:::   |||   |||   |||:::         :::
 651 luSerGlyTyrAspLys...PheLysGlyArgThrTyrTrpAsnLeuThr 666

 961 TTCTACGAACCACGTCAAAATGGGAAATACTCTTTTAACGACGATAATAA 1010
           :::      ::::::|||:::::::::::::::   |||:::::
 667 SerLeuAsnValSerGluSerGlyGluPheAsnLeuThrIleAspSerAr 683

1011 TGGCACA.........GGAAAAATCAATGCCAAACATGAACACAATTCTC 1051
     :|||:::         |||   ::::::      ::::::   ||::::
 683 gGlySerAspSerAlaGlyThrLeuThrGlnProTyrAsnLeuAsnGlyI 700

1052 TGCCT...AATAGATTAAAAACACGAACCGTTCAATTGTTTAATGTTTCT 1098
     ::      |||:::     |||   :::|||:::      |||:::
 700 leSerPheAsnLysAspThrThrPheAsnValGluArg...AsnAlaArg 715

1099 TTATCCGAGACAGCAAGAGAACCTGTTTATCATGCTGCAGGTGGTGTCAA 1148
     ::::::          :::   |||:::               |||:::||
 716 ValAsnPheAspIleLysAlaProIle.............GlyIleAs 727

1149 CAGTTATCGACCCAGACTGAATAATGGAGAAAATATTTCCTTTATTGACG 1198
     |:::|||                        ::::::::::::::   :::
 727 nLysTyr.....................SerSerLeuAsnTyrAlaSer. 736

1199 AAGGAAAAGGCGAATTGATACTTACCAGCAACATCAATCAAGGTGCTGGA 1248
                        :::::::::||||||:::   ::::::|||
 737 ...................PheAsnGlyAsnIleSerValSerGlyGly 746

1249 GGATTATATTTCCAAGGAGATTTTACGGTCTCGCCTGAAAATAACGAAAC 1298
     |||          :::   |||||||||:::       ::::::::::::
 747 Gly.........SerValAspPheThrLeuLeuAlaSerSerSerAsnVa 760

1299 TTGGCAAGGCGCGGGCGTT.....................CATATCAGTG 1327
     :   |||      ||||||                      ::::::|||
 760 l...GlnThrProGlyValValIleAsnSerLysTyrPheAsnValSerT 776

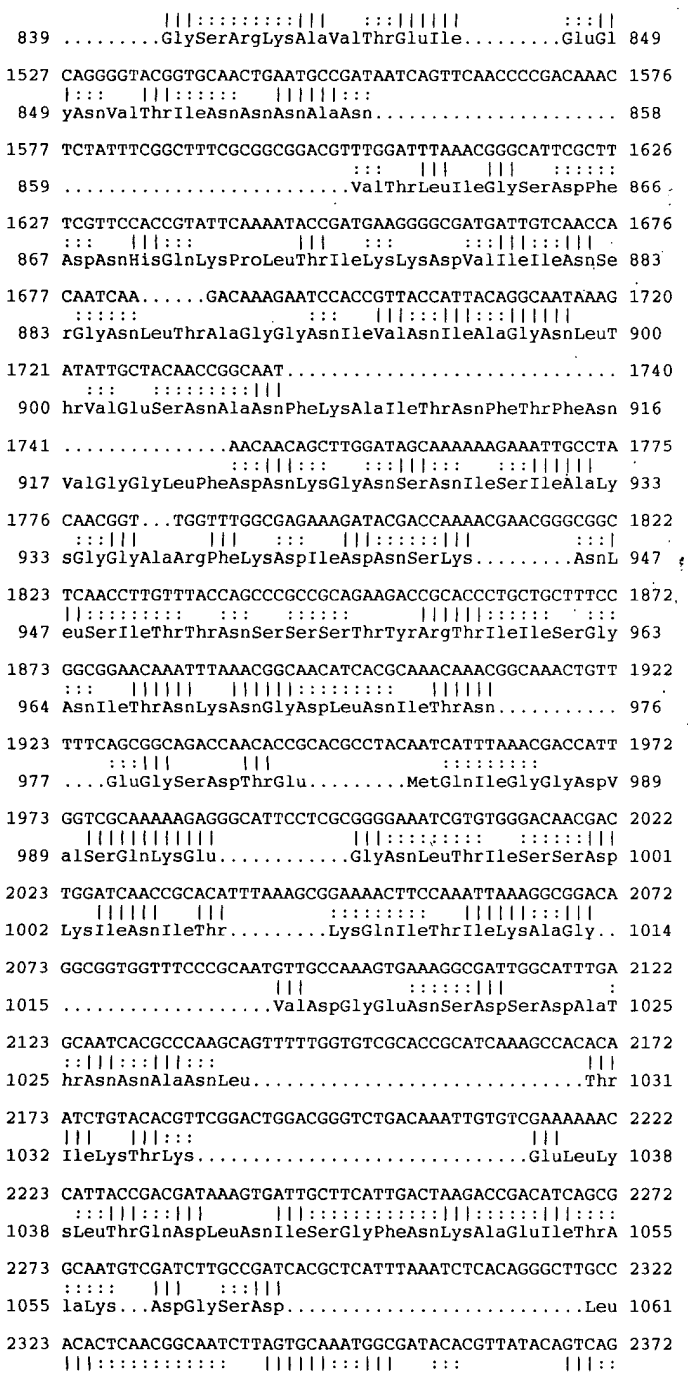
1328 AAGACAGTACCGTTACTTGGAAAGTAAACGGC..................... 1359
          |||::::::   :::|||::::::|||
 776 hrGlySerSerLeuArgPheLysThrSerGlySerThrLysThrGlyPhe 792

1360 ...GTGGCAAACGAC...CGCCTGTCCAAAATCGGCAAAGGCACGCTGCA 1403
        :::   :::|||        |||:::      |||   :::
 793 SerIleGluLysAspLeuThrLeuAsnAlaThrGlyGlyAsnIleThrLe 809

1404 CGTTCAAGCCAAAGGGGAAAAC..........................C 1426
      :::|||::::::|||   :::                          :
 809 uLeuGlnValGluGlyThrAspGlyMetIleGlyLysGlyIleValAlaL 826

1427 AAGGCTCGATCAGCGTGGGCGACGGTACAGTCATTTTGGATCAGCAGGCA 1476
     ::   :::|||:::         |||::::::   :::
 826 ysLysAsnIleThrPheGluGlyGlyAsnIleThrPhe............ 838

1477 GACGATAAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCTTGGTCAGCGG 1526
```


839 GlySerArgLysAlaValThrGluIle.....GluGl 849
1527 CAGGGGTACGGTGCAACTGAATGCCGATAATCAGTTCAACCCCGACAAAC 1576
849 yAsnValThrIleAsnAsnAsnAlaAsn..... 858
1577 TCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAACGGGCATTCGCTT 1626
859 ValThrLeuIleGlySerAspPhe 866
1627 TCGTTCCACCGTATTCAAAATACCGATGAAGGGGCGATGATTGTCAACCA 1676
867 AspAsnHisGlnLysProLeuThrIleLysLysAspValIleIleAsnSe 883
1677 CAATCAA......GACAAAGAATCCACCGTTACCATTACAGGCAATAAAG 1720
883 rGlyAsnLeuThrAlaGlyGlyAsnIleValAsnIleAlaGlyAsnLeuT 900
1721 ATATTGCTACAACCGGCAAT..... 1740
900 hrValGluSerAsnAlaAsnPheLysAlaIleThrAsnPheThrPheAsn 916
1741 AACAACAGCTTGGATAGCAAAAAAGAAATTGCCTA 1775
917 ValGlyGlyLeuPheAspAsnLysGlyAsnSerAsnIleSerIleAlaLy 933
1776 CAACGGT...TGGTTTGGCGAGAAAGATACGACCAAAACGAACGGGCGGC 1822
933 sGlyGlyAlaArgPheLysAspIleAspAsnSerLys.....AsnL 947
1823 TCAACCTTGTTTACCAGCCCGCCGCAGAAGACCGCACCCTGCTGCTTTCC 1872
947 euSerIleThrThrAsnSerSerSerThrTyrArgThrIleIleSerGly 963
1873 GGCGGAACAAATTTAAACGGCAACATCACGCAAACAAACGGCAAACTGTT 1922
964 AsnIleThrAsnLysAsnGlyAspLeuAsnIleThrAsn..... 976
1923 TTTCAGCGGCAGACCAACACCGCACGCCTACAATCATTTAAACGACCATT 1972
977 GluGlySerAspThrGlu.....MetGlnIleGlyGlyAspV 989
1973 GGTCGCAAAAAGAGGGCATTCCTCGCGGGGAAATCGTGTGGGACAACGAC 2022
989 alSerGlnLysGlu.....GlyAsnLeuThrIleSerSerAsp 1001
2023 TGGATCAACCGCACATTTAAAGCGGAAAACTTCCAAATTAAAGGCGGACA 2072
1002 LysIleAsnIleThr.....LysGlnIleThrIleLysAlaGly.. 1014
2073 GGCGGTGGTTTCCCGCAATGTTGCCAAAGTGAAAGGCGATTGGCATTTGA 2122
1015 ValAspGlyGluAsnSerAspSerAspAlaT 1025
2123 GCAATCACGCCCAAGCAGTTTTTGGTGTCGCACCGCATCAAAGCCACACA 2172
1025 hrAsnAsnAlaAsnLeu.....Thr 1031
2173 ATCTGTACACGTTCGGACTGGACGGGTCTGACAAATTGTGTCGAAAAAAC 2222
1032 IleLysThrLys.....GluLeuLy 1038
2223 CATTACCGACGATAAAGTGATTGCTTCATTGACTAAGACCGACATCAGCG 2272
1038 sLeuThrGlnAspLeuAsnIleSerGlyPheAsnLysAlaGluIleThrA 1055
2273 GCAATGTCGATCTTGCCGATCACGCTCATTTAAATCTCACAGGGCTTGCC 2322
1055 laLys...AspGlySerAsp.....Leu 1061
2323 ACACTCAACGGCAATCTTAGTGCAAATGGCGATACACGTTATACAGTCAG 2372
1062 ThrIleGlyAsnThrAsnSerAlaAspGlyThrAsnAlaLysLysValTh 1078
2373 CCACAACGCCACCCAAAACGGCAACCTTAGCCTCGTGGGCAATGCCCAAG 2422
1078 rPheAsnGlnValLysAspSerLysIleSerAlaAspGly..... 1091
2423 CAACATTTAATCAAGCCACATTAAACGGCAACACATCGGCTTCGGGCAAT 2472
1092 HisLysValThrLeuHisSerLysValGluThrSerGlySer 1105
2473 GCTTCATTTAATCTAAGCGACCACGCCGTACAAAACGGCAGTCTGACGCT 2522
1106 AsnAsn...AsnThrGluAspSerSerAspAsnAsnAlaGlyLeuThrIl 1121
2523 TTCCGGCAACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACGGTAATG 2572
1121 e.....AspAlaLys...AsnValThr.....ValAsnAsnAsnI 1132
2573 TCTCCCTAGCCGATAAGGCAGTATTCCATTTTGAAAGCAGCCGCTTTACC 2622
1132 leThr...SerHisLysAlaVal.....SerIleSerAlaThrSer 1144
2623 GGACAAATCAGCGGCGGCAAGGATACGGCATTACACTTAAAA.....GA 2666
1145 GlyGluIleThrThrLysThrGlyThrThrIleAsnAlaThrThrGlyAs 1161
2667 CAGCGAATGGACGCTGCCGTCAGGCACGGAATTAGGCAATTTAAACCTTG 2716
1161 nValGluIleThrAlaGlnThrGlySerIleLeuGlyGlyIleGluSerS 1178
2717 ACAACGCCACCATTACACTCAATTCCGCCTATCGCCACGATGCGGCAGGG 2766
1178 erSerGlySerValThrLeuThrAlaThrGluGlyAlaLeuAlaValSer 1194
2767 GCGCAAACCGGCAGTGCGACAGATGCGCCGCGCCGCCGTTCGCGCCGTTC 2816
1195 AsnIleSerGlyAsnThrValThrValThrAlaAsnSerGlyAlaLeuTh 1211
2817 GCGCCGTTCCCTATTATCCGTTACACCGCCAACTTCGGTAGAATCCCGTT 2866
1211 rThrLeuAlaGlySerThrIleLysGlyThrGluSerValThrThrSerS 1228
2867 TCAACACGCTGACGGTAAACGGCAAATTGAACGGTCAGGGAACA..... 2910
1228 erGlnSerGlyAspIleGlyGlyThrIleSerGly...GlyThrValGlu 1243
2911 TTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGACAAATTGAAGCT 2960
1244 ValLysAlaThrGluSerLeuThrThrGlnSerAsnSerLysIleLys.. 1259
2961 GGCGGAAAGTTCCGAAGGCACTTACACCTTGGCGGTC..... 2997
1260 .AlaThrThrGlyGluAlaAsnValThrSerAlaThrGlyThrIleGlyG 1276
2998 AACAATACCGGCAACGAACCTGCAAGCCTCGAACAATTG 3036
1276 lyThrIleSerGlyAsnThrValAsnValThrAlaAsnAlaGlyAspLeu 1292
3037 ACGGTA.....GTGGAAGGAAAAGACAACAA 3062
1293 ThrValGlyAsnGlyAlaGluIleAsnAlaThrGluGlyAlaAlaThrLe 1309
3063 ACCGCTGTCCGAAAACCTTAATTTCACCCTGCAAAACGAACACGTCGATG 3112
1309 uThrThrSerSerGlyLysLeuThrThrGluAlaSerSerHisIleThrS 1326
3113 CCGGCGCGTGGCGTTACCAACTCATCCGCAAAGACGGCGAGTTCCGCCTG 3162
1326 er.AlaLysGlyGlnValAsnLeuSerAlaGlnAspGlySerValAlaGl 1342
3163 CATAATCCGGTCAAAGAACAAGAGCTTTCCGACAAACTCGGCAAGGCAGA 3212
1342 ySerIleAsnAlaAlaAsnValThrLeuAsnThrThrGlyThrLeuThrT 1359

```
    3213 AGCCAAAAAACAGGCGGAAAAAGACAACGCGCAAAGCCTTG 3253
    1359 hrValLysGlySerAsnIleAsnAlaThrSerGlyThrLeu 1372

seq_name: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-719-641-2

seq_documentation_block:
; Sequence 2, Application US/08719641
; Patent No. 6218141
;  GENERAL INFORMATION:
;    APPLICANT:  Barenkamp, Stephen J
;    TITLE OF INVENTION:  High Molecular Weight Surface Proteins
;    TITLE OF INVENTION:  of No. 6218141-Typeable Haemophilus
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Shoemaker and Mattare, Ltd.
;      STREET:  2001 Jefferson Davis Hwy., 1203 Crystal Plaza
;      STREET:  Bldg. 1
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202-0286
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/719,641
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/302,832
;      FILING DATE:  16-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US PCT/US93/02166
;      FILING DATE:  16-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9205704.1
;      FILING DATE:  16-MAR-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Berkstresser, Jerry W
;      REGISTRATION NUMBER:  22,651
;      REFERENCE/DOCKET NUMBER:  1038-625
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 415-0810
;      TELEFAX:  (703) 415-0813
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1536 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:
US-08-719-641-2


alignment_scores:
            Quality:  272.50            Length:   1182
              Ratio:  0.452               Gaps:     60
 Percent Similarity:  51.015   Percent Identity: 20.897

alignment_block:
 US-09-303-518D-649 x US-08-719-641-2   ..

 Align seg 1/1  to: US-08-719-641-2  from: 1  to: 1536

     127 GGACACACTTATTTCGGCATCAACTACCAATACTATCGCGACTTTGCCGA 176
     349 GlyGluThrTyrLeuGly.........................GlyAs 356

     177 AAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACA 226

     356 pGluArgGlyGlu.....GlyLysAsnGlyIleGlnLeuAlaLysL 370

     227 AAAAAGGGGAGTTGGTCGGCAAATCAATGACAAAAGCCCCGATGATTGAT 276
     370 ys..................ThrSerLeuGluLysGlySerThrIleAsn 380

     277 TTTTCTGTGGTGTCGCGTAACGGCGTGGCGGCATTGGTGGGCGATCAATA 326
     381 ValSerGlyLysGluLysGlyGlyArgAlaIleValTrpGlyAspIleAl 397

     327 TATTGTG.............................AGCGTGGCACATA 346
     397 aLeuIleAspGlyAsnIleAsnAlaGlnGlySerGlyAspIleAlaLysT 414

     347 ACGGCGGC..................................... 354
     414 hrGlyGlyPheValGluThrSerGlyHisAspLeuPheIleLysAspAsn 430

     355 ...............................TATAACAACGTTGATTT 371
     431 AlaIleValAspAlaLysGluTrpLeuLeuAspPheAspAsnValSerIl 447

     372 TGGTGCGGAA......GGAAGAAATCCCGATCAACATCGTTTTACTTAT. 414
     447 eAsnAlaGluThrAlaGlyArgSerAsnThrSerGluAspAspGluTyrT 464

     415 ...................AAAATTGTGAAACGGAATAATTATAAAGCA 444
     464 hrGlySerGlyAsnSerAlaSerThrProLysArgAsnLysGluLysThr 480

     445 GGGACTAAAGGCCATCCTTATGGCGGCGATTATCATATGCCGCGTTTGCA 494
     481 .ThrLeuThrAsnThrThrLeuGluSerIleLeuLysLysGlyThrPheV 497

     495 TAAATTTGTCACAGATGCAGAACCTGTTGAAATGACCAGTTATATGGATG 544
     497 alAsnIleThrAlaAsnGlnArg..................IleTyrVal 507

     545 GGCGGAAATATATCGATCAAAATAATTACCCTGACCGTGTTCGTATTGGG 594
     508 AsnSerSerIleAsnLeuSerAsnGlySerLeuThrLeuTrpSerGluGl 524

     595 GCAGGCAGGC..........................AATATTGGCGATC 617
     524 yArgSerGlyGlyGlyValGluIleAsnAsnAspIleThrThrGlyAspA 541

     618 TGATGAAGATGAGCCCAATAACC..................GCGAAAGTT 649
     541 spThrArgGlyAlaAsnLeuThrIleTyrSerGlyGlyTrpValAspVal 557

     650 CATATCATATTGCAAGTGCGTATTCTTGGCTCGTTGGTGGCAATACCTTT 699
     558 HisLysAsnIleSerLeuGlyAlaGlnGlyAsnIle..AsnIleThrAla 573

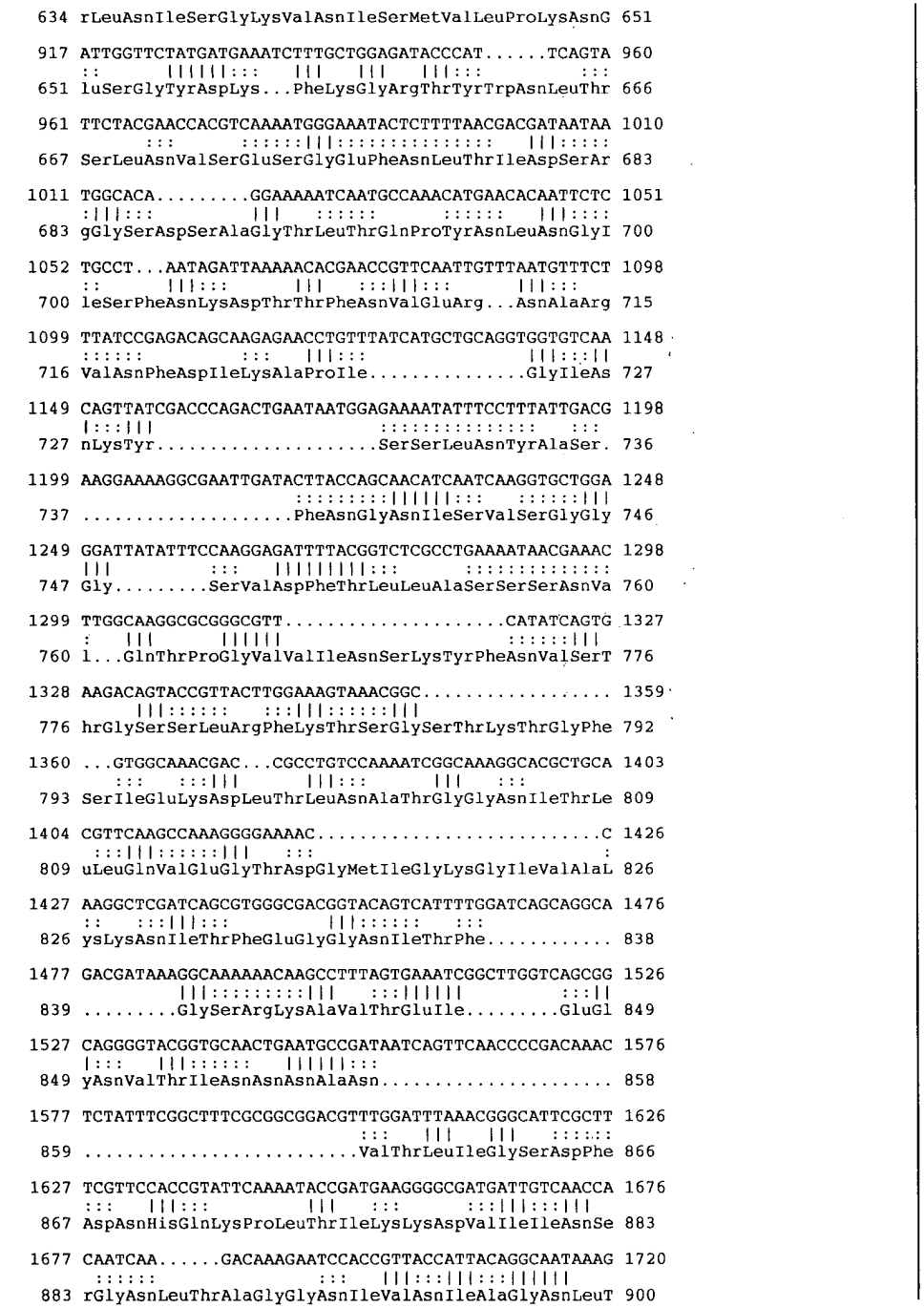
     700 GCACAAAAT..............GGATCA..............GGTGG 719
     574 LysGlnAspIleAlaPheGluLysGlySerAsnGlnValIleThrGlyGl 590

     720 TGGCACAGTCAACTTAGGTAGTGAAAAAATTAAACATAGCCCATATGGTT 769
     590 nGlyThrIleThrSerGlyAsnGlnLys..................GlyP 601

     770 TTTTACCAACAGGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATC 819
     601 heArgPheAsnAsnValSerLeuAsnGlyThrGlySerGlyLeuGlnPhe 617

     820 TATGATGCCCAAAAGCAAAAGTGGTTAATTAATGGGGTATTGCAAACGGG 869
     618 ThrThrLysArgThrAsnLysTyrAlaIleThrAsnLysPheGluGlyTh 634

     870 CAACCCCTATATAGGAAAAAGCAAT...GGCTTCCAGCTGGTTCGTAAAG 916
```


634 rLeuAsnIleSerGlyLysValAsnIleSerMetValLeuProLysAsnG 651
917 ATTGGTTCTATGATGAAATCTTTGCTGGAGATACCCAT......TCAGTA 960
651 luSerGlyTyrAspLys...PheLysGlyArgThrTyrTrpAsnLeuThr 666
961 TTCTACGAACCACGTCAAAATGGGAAATACTCTTTTAACGACGATAATAA 1010
667 SerLeuAsnValSerGluSerGlyGluPheAsnLeuThrIleAspSerAr 683
1011 TGGCACA.....GGAAAAATCAATGCCAAACATGAACACAATTCTC 1051
683 gGlySerAspSerAlaGlyThrLeuThrGlnProTyrAsnLeuAsnGlyI 700
1052 TGCCT...AATAGATTAAAAACACGAACCGTTCAATTGTTTAATGTTTCT 1098
700 leSerPheAsnLysAspThrThrPheAsnValGluArg...AsnAlaArg 715
1099 TTATCCGAGACAGCAAGAGAACCTGTTTATCATGCTGCAGGTGGTGTCAA 1148
716 ValAsnPheAspIleLysAlaProIle.....GlyIleAs 727
1149 CAGTTATCGACCCAGACTGAATAATGGAGAAAATATTTCCTTTATTGACG 1198
727 nLysTyr.....SerSerLeuAsnTyrAlaSer. 736
1199 AAGGAAAAGGCGAATTGATACTTACCAGCAACATCAATCAAGGTGCTGGA 1248
737 PheAsnGlyAsnIleSerValSerGlyGly 746
1249 GGATTATATTTCCAAGGAGATTTTACGGTCTCGCCTGAAAATAACGAAAC 1298
747 Gly.....SerValAspPheThrLeuLeuAlaSerSerSerAsnVa 760
1299 TTGGCAAGGCGCGGGCGTT.....CATATCAGTG 1327
760 l...GlnThrProGlyValValIleAsnSerLysTyrPheAsnValSerT 776
1328 AAGACAGTACCGTTACTTGGAAAGTAAACGGC..... 1359
776 hrGlySerSerLeuArgPheLysThrSerGlySerThrLysThrGlyPhe 792
1360 ...GTGGCAAACGAC...CGCCTGTCCAAAATCGGCAAAGGCACGCTGCA 1403
793 SerIleGluLysAspLeuThrLeuAsnAlaThrGlyGlyAsnIleThrLe 809
1404 CGTTCAAGCCAAAGGGGAAAAC.....C 1426
809 uLeuGlnValGluGlyThrAspGlyMetIleGlyLysGlyIleValAlaL 826
1427 AAGGCTCGATCAGCGTGGGCGACGGTACAGTCATTTTGGATCAGCAGGCA 1476
826 ysLysAsnIleThrPheGluGlyGlyAsnIleThrPhe..... 838
1477 GACGATAAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCTTGGTCAGCGG 1526
839 GlySerArgLysAlaValThrGluIle.....GluGl 849
1527 CAGGGGTACGGTGCAACTGAATGCCGATAATCAGTTCAACCCCGACAAAC 1576
849 yAsnValThrIleAsnAsnAsnAlaAsn..... 858
1577 TCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAACGGGCATTCGCTT 1626
859 ValThrLeuIleGlySerAspPhe 866
1627 TCGTTCCACCGTATTCAAAATACCGATGAAGGGGCGATGATTGTCAACCA 1676
867 AspAsnHisGlnLysProLeuThrIleLysLysAspValIleIleAsnSe 883
1677 CAATCAA......GACAAAGAATCCACCGTTACCATTACAGGCAATAAAG 1720
883 rGlyAsnLeuThrAlaGlyGlyAsnIleValAsnIleAlaGlyAsnLeuT 900
1721 ATATTGCTACAACCGGCAAT..... 1740
900 hrValGluSerAsnAlaAsnPheLysAlaIleThrAsnPheThrPheAsn 916
1741 AACAACAGCTTGGATAGCAAAAAAGAAATTGCCTA 1775
917 ValGlyGlyLeuPheAspAsnLysGlyAsnSerAsnIleSerIleAlaLy 933
1776 CAACGGT...TGGTTTGGCGAGAAAGATACGACCAAAACGAACGGGCGGC 1822
933 sGlyGlyAlaArgPheLysAspIleAspAsnSerLys.....AsnL 947
1823 TCAACCTTGTTTACCAGCCCGCCGCAGAAGACCGCACCCTGCTGCTTTCC 1872
947 euSerIleThrThrAsnSerSerSerThrTyrArgThrIleIleSerGly 963
1873 GGCGGAACAAATTTAAACGGCAACATCACGCAAACAAACGGCAAACTGTT 1922
964 AsnIleThrAsnLysAsnGlyAspLeuAsnIleThrAsn..... 976
1923 TTTCAGCGGCAGACCAACACCGCACGCCTACAATCATTTAAACGACCATT 1972
977 GluGlySerAspThrGlu.....MetGlnIleGlyGlyAspV 989
1973 GGTCGCAAAAAGAGGGCATTCCTCGCGGGGAAATCGTGTGGGACAACGAC 2022
989 alSerGlnLysGlu.....GlyAsnLeuThrIleSerSerAsp 1001
2023 TGGATCAACCGCACATTTAAAGCGGAAAACTTCCAAATTAAAGGCGGACA 2072
1002 LysIleAsnIleThr.....LysGlnIleThrIleLysAlaGly.. 1014
2073 GGCGGTGGTTTCCCGCAATGTTGCCAAAGTGAAAGGCGATTGGCATTTGA 2122
1015 ValAspGlyGluAsnSerAspSerAspAlaT 1025
2123 GCAATCACGCCCAAGCAGTTTTTGGTGTCGCACCGCATCAAAGCCACACA 2172
1025 hrAsnAsnAlaAsnLeu.....Thr 1031
2173 ATCTGTACACGTTCGGACTGGACGGGTCTGACAAATTGTGTCGAAAAAAC 2222
1032 IleLysThrLys.....GluLeuLy 1038
2223 CATTACCGACGATAAAGTGATTGCTTCATTGACTAAGACCGACATCAGCG 2272
1038 sLeuThrGlnAspLeuAsnIleSerGlyPheAsnLysAlaGluIleThrA 1055
2273 GCAATGTCGATCTTGCCGATCACGCTCATTTAAATCTCACAGGGCTTGCC 2322
1055 laLys...AspGlySerAsp.....Leu 1061
2323 ACACTCAACGGCAATCTTAGTGCAAATGGCGATACACGTTATACAGTCAG 2372
1062 ThrIleGlyAsnThrAsnSerAlaAspGlyThrAsnAlaLysLysValTh 1078
2373 CCACAACGCCACCCAAAACGGCAACCTTAGCCTCGTGGGCAATGCCCAAG 2422
1078 rPheAsnGlnValLysAspSerLysIleSerAlaAspGly..... 1091
2423 CAACATTTAATCAAGCCACATTAAACGGCAACACATCGGCTTCGGGCAAT 2472
1092 HisLysValThrLeuHisSerLysValGluThrSerGlySer 1105
2473 GCTTCATTTAATCTAAGCGACCACGCCGTACAAAACGGCAGTCTGACGCT 2522
1106 AsnAsn...AsnThrGluAspSerSerAspAsnAsnAlaGlyLeuThrIl 1121
2523 TTCCGGCAACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACGGTAATG 2572
1121 e.....AspAlaLys...AsnValThr.....ValAsnAsnAsnI 1132

```
     2573 TCTCCCTAGCCGATAAGGCAGTATTCCATTTTGAAAGCAGCCGCTTTACC 2622
          :::::   :::   |||||||||         :::   |||      :::
     1132 leThr...SerHisLysAlaVal.........SerIleSerAlaThrSer 1144

     2623 GGACAAATCAGCGGCGGCAAGGATACGGCATTACACTTAAAA......GA 2666
          |||:::|||:::           |||:::::::::            ::
     1145 GlyGluIleThrThrLysThrGlyThrThrIleAsnAlaThrThrGlyAs 1161

     2667 CAGCGAATGGACGCTGCCGTCAGGCACGGAATTAGGCAATTTAAACCTTG 2716
          :   |||   |||      :::|||:::   ||||||:::::::::   .:
     1161 nValGluIleThrAlaGlnThrGlySerIleLeuGlyGlyIleGluSerS 1178

     2717 ACAACGCCACCATTACACTCAATTCCGCCTATCGCCACGATGCGGCAGGG 2766
          ::::::::::::::|||||:::::::::           |||::::::
     1178 erSerGlySerValThrLeuThrAlaThrGluGlyAlaLeuAlaValSer 1194

     2767 GCGCAAACCGGCAGTGCGACAGATGCGCCGCGCCGCCGTTCGCGCCGTTC 2816
               :::|||:::::::::   :::      :::   :::      ::
     1195 AsnIleSerGlyAsnThrValThrValThrAlaAsnSerGlyAlaLeuTh 1211

     2817 GCGCCGTTCCCTATTATCCGTTACACCGCCAACTTCGGTAGAATCCCGTT 2866
          :      :::      ::::::           ||||||   :::
     1211 rThrLeuAlaGlySerThrIleLysGlyThrGluSerValThrThrSerS 1228

     2867 TCAACACGCTGACGGTAAACGGCAAATTGAACGGTCAGGGAACA..... 2910
            ::::::      :::::::|||   :::::::|||   ||||||
     1228 erGlnSerGlyAspIleGlyGlyThrIleSerGly...GlyThrValGlu 1243

     2911 TTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGACAAATTGAAGCT 2960
            :::      ::::::|||            ::::::|||:::|||
     1244 ValLysAlaThrGluSerLeuThrThrGlnSerAsnSerLysIleLys.. 1259

     2961 GGCGGAAAGTTCCGAAGGCACTTACACCTTGGCGGTC............. 2997
           |||   ::::::|||::::::   |||   |||:::
     1260 .AlaThrThrGlyGluAlaAsnValThrSerAlaThrGlyThrIleGlyG 1276

     2998 ..........AACAATACCGGCAACGAACCTGCAAGCCTCGAACAATTG 3036
                    :::||||||   |||      |||:::      :::|||
     1276 lyThrIleSerGlyAsnThrValAsnValThrAlaAsnAlaGlyAspLeu 1292

     3037 ACGGTA......................GTGGAAGGAAAAGACAACAA 3062
          ||||||                     :::||||||      :::
     1293 ThrValGlyAsnGlyAlaGluIleAsnAlaThrGluGlyAlaAlaThrLe 1309

     3063 ACCGCTGTCCGAAAACCTTAATTTCACCCTGCAAAACGAACACGTCGATG 3112
               |||::::::         |||      ::::::|||:::   :
     1309 uThrThrSerSerGlyLysLeuThrThrGluAlaSerSerHisIleThrS 1326

     3113 CCGGCGCGTGGCGTTACCAACTCATCCGCAAAGACGGCGAGTTCCGCCTG 3162
          :: |||:::|||   :::|||   |||||||:::   :::|||   |||
     1326 er.AlaLysGlyGlnValAsnLeuSerAlaGlnAspGlySerValAlaGl 1342

     3163 CATAATCCGGTCAAAGAACAAGAGCTTTCCGACAAACTCGGCAAGGCAGA 3212
               |||::::::   |||   ::::::   |||:::::::::
     1342 ySerIleAsnAlaAlaAsnValThrLeuAsnThrThrGlyThrLeuThrT 1359

     3213 AGCCAAAAAACAGGCGGAAAAAGACAACGCGCAAAGCCTTG 3253
               |||:::   :::   :::::::|||      :::|||
     1359 hrValLysGlySerAsnIleAsnAlaThrSerGlyThrLeu 1372

seq_name: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-169-927-2

seq_documentation_block:
; Sequence 2, Application US/08169927
; Patent No. 5783441
;  GENERAL INFORMATION:
;    APPLICANT:  Carl, Mitchell
;    APPLICANT:  Dobson, Michael E.
;    APPLICANT:  Ching, Wei Mei
;    APPLICANT:  Dasch, Gregory A
;    TITLE OF INVENTION:  Gene and Protein Applicable to the
;    TITLE OF INVENTION:  Preparation of Vaccines for Rickettsia prowazekii and
;    TITLE OF INVENTION:  Rickettsia typhi and the Detection of Both
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Counsel, Naval Medical R & D Command
;      STREET:  Bldg. 1, T-12, 8901 Wisconsin Ave.
;      CITY:  Bethesda
;      STATE:  MD
;      COUNTRY:  USA
;      ZIP:  20889-5606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/169,927
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/742,128
;      FILING DATE:  08/09/91
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Spevack, A. David
;      REGISTRATION NUMBER:  24,743
;      REFERENCE/DOCKET NUMBER:  75,976
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 295-6759
;      TELEFAX:  (301) 295-1022
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1612 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-169-927-2


alignment_scores:
            Quality:  241.00              Length:   1485
              Ratio:  0.361                 Gaps:     75
 Percent Similarity:  44.916   Percent Identity: 19.259

alignment_block:
 US-09-303-518D-649 x US-08-169-927-2   ..

 Align seq 1/1  to: US-08-169-927-2  from: 1  to: 1612

      634 AATAACCGCGAAAGTTCATATCATATTGCAAGTGCGTATTCTTGGCTCGT 683
          |||:::|||:::|||||||||||||:::::|||
      160 AsnSerArgSerSerSerTyrHisLeuValSer..................... 170

      684 TGGTGGCAATACCTTTGCACAAAATGGATCAGGTGGTGGCACAGTCAACT 733
                ||||::         ||||||   |||:::|||   :::|||:
      171 .......AsnSerLysIleValAsnGly...GlyAsnGlyIleLeuAsnI 184

      734 TA......GGTAGTGAAAAAATTAAACATAGCCCATATGGTTTTTTACCA 777
          ::      |||      :::::::::   :::   ::::::   :::
      184 leThrAsnGlyPheIleGlnValSerAspAsnThrPheAlaGlyIleLys 200

      778 ACAGGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATGC 827
          |||       ::::::   |||       :::   ||||||
      201 Thr...IleAsnIleAspAspCysGlnGlyLeuMetPheAsnSerThrPr 216

      828 CCAAAAGCAAAAGTGGTTA.............................A 847
           :::      :::   |||                            :
      216 oAspAlaAlaAsnThrLeuAsnLeuGlnAlaGlyGlyAsnThrIleAsnP 233

      848 TTAATGGGGTATTGCAAACGGGCAACCCCTATATAGGAAAAAGCAATGGC 897
          ::|||||:::      ||||||:::      :::   ::::::||||||
      233 heAsnGlyIleAspGlyThrGlyLysLeuValLeuValSerLysAsnGly 249

      898 ............TTCCAGCTGGTTCGTAAAGATTGGTTCTATGATGAAAT 935
                      |||:::::::::       :::|||:::   :::|||:
      250 AlaAlaThrGluPheAsnValThrGly.ThrLeuGlyGlyAsnLeuLysG 266
```



```
 765 eSerAsnAsnTyrThrThrAspHisValGluSerAlaAspAsnThrGlyT 782

2396 ACCTTAGCCTCGTGGGCAATGCCCAAGCAACATTTAATCAAGCCACATTA 2445
 782 hrLeuGluPheValAsnThrAspProIleThr.....ValThrLeu 795

2446 AACGGCAAC.............................ACATCGGC 2462
 796 AsnLysGlnGlyAlaTyrPheGlyValLeuLysGlnValIleIleSerGl 812

2463 TTCGGGCAATGCTTCATTTAAT.......................... 2484
 812 yProGlyAsnIleValPheAsnGluIleGlyAsnValGlyIleValHisG 829

2485 ........CTAAGCGACCACGCCGTACAAAACGGCAGTCTGACGCTTTCC 2526
 829 lyIleAlaAlaAsnSerIleSerPheGluAsnAlaSerLeuGlyThrSer 845

2527 ..............................GGCAACGCTAAGGCAAA 2543
 846 LeuPheLeuProSerGlyThrProLeuAspValLeuThrIleLysSerTh 862

2544 CGTAAGCCATTCCGCACTCAACGGTAATGTCTCCCTAGCCGATAAGGCAG 2593
 862 rValGlyAsnGlyThrValAsp...AsnPheAsnAlaProIleValValV 878

2594 TATTCCATTTTGAAAGCAGCCGCTTTACCGGACAAATCAGCGGCGGCAAG 2643
 878 alSerGlyIleAspSerMetIleAsnAsnGlyGlnIleIleGlyAspLys 894

2644 GATACGGCATTACACTTA.........AAAGACAGCGAATGGACGCTGCC 2684
 895 LysAsnIleIleAlaLeuSerLeuGlySerAspAsnSerIleThrValAs 911

2685 GTCAGGCACGGAATTAGGCAATTTAAACCTT.........GACAACGCCA 2725
 911 nAlaAsnThrLeuTyrSerGlyIleArgThrThrLysAsnAsnGlnGlyT 928

2726 CCATTACACTCAATTCCGCCTATCGCCACGAT..............GCG 2760
 928 hrValThrLeuSerGlyGlyMetProAsnAsnProGlyThrIleTyrGly 944

2761 GCAGGGGCGCAAACCGGCAGTGCGACAGATGCGCCGCGCCGCCGTTCGCG 2810
 945 LeuGlyLeuGluAsnGlySerPro........................... 952

2811 CCGTTCGCGCCGTTCCCTATTATCCGTTACACCGCCAACTTCGGTAGAAT 2860
 953 .............LysLeuLysGlnValThrPheThrThrAspTyrAsnA 965

2861 CCCGT............TTCAACACGCTGACGGTAAAC............. 2886
 965 snLeuGlySerIleIleAlaAsnAsnValThrIleAsnAspAspValThr 981

2886 ................................................... 2886
 982 LeuThrThrGlyGlyIleAlaGlyThrAspPheAspAlaLysIleThrLe 998

2887 .GGCAAATTGAACGGTCAGGGAACATTCCGCTTTATGTCGGAACTCTTCG 2935
 998 uGlySerValAsnGlyAsnAlaAsnValArgPheValAspSerThrPheS 1015

2936 GCTACCGCAGCGACAAATTGAAGCTGGCGGAAAGTTCCGAAGGCACTTAC 2985
1015 erAspProArgSerMetIleValAlaThrGlnAlaAsnLysGlyThrVal 1031

2986 ACC............TTGGCGGTCAACAATACCGGC......AACGAACC 3017
1032 ThrTyrLeuGlyAsnAlaLeuValSerAsnIleGlySerLeuAspThrPr 1048

3018 TGCAAGCCTCGAACAATTGACGGTAGTGGAAGGAAAAGACAACAAACCGC 3067
```

```
1048 oValAlaSerValArgPheThr.........GlyAsnAspSer....... 1059

3068 TGTCCGAAAACCTTAATTTCACCCTGCAAAACGAACACGTCGATGCCGGC 3117
1060 ..GlyAlaGlyLeuGlnGlyAsnIleTyrSerGlnAsnIleAspPheGly 1075

3118 GCGTGGCGTTACCAACTCATCCGCAAAGACGGCGAGTTCCGCCTGCATAA 3167
1076 ThrTyrAsnLeuThrIleLeuAsnSer.....................As 1085

3168 TCCGGTCAAAGAACAAGAGCTTTCCGACAAACTCGGCAAGGCAGAAGCCA 3217
1085 nValIle.......................LeuGlyGlyGlyThrThrA 1094

3218 AAAAACAGGCGGAAAAAGACAACGCGCAAAGCCTTGACGCGCTGATTGCG 3267
1094 laIleAsnGlyGluIleAspLeuLeuThrAsn...AsnLeuIlePheAla 1109

3268 GCCGGGCGCGATGCCGTCGAAAAGACAGAAAGCGTTGCCGAACCGGCCCG 3317
1110 AsnGlyThrSerThrTrpGlyAspAsnThrSerIleSerThrThrLeuAs 1126

3318 GCAGGCAGGCGGGGAAAATGTCGGC...ATTATGCAGGCGGAGGAAGAGA 3364
1126 nValSerSerGly...AsnIleGlyGlnValValIleAlaGluAspAlaG 1142

3365 AAAAACGGGTGCAGGCGGATAAAGACACCGCCTTGGCGAAACAGCGCGAA 3414
1142 lnValAsnAlaThrThrThrGlyThrThrThrIleLysIleGlnAspAsn 1158

3415 GCGGAAACCCGG...CCGGCTACCACCGCCTTCCCCCGCGCCCGCCGCGC 3461
1159 AlaAsnAlaAsnPheSerGlyThrGlnAlaTyrThrLeuIleGlnGlyGl 1175

3462 CCGCCGGGATTTGCCGCAACTGCAACCCCAACCGCAGCCCCAACCGCAGC 3511
1175 yAlaArgPheAsnGlyThrLeu........................... 1182

3512 GCGACCTGATCAGCCGTTATGCCAATAGCGGTTTGAGTGAATTTTCCGCC 3561
1183 ............................GlyAlaProAsnPheAlaVal 1189

3562 ACGCTCAACAGCGTTTTCGCCGTACAGGACGAATTAGACCGCGTATTTGC 3611
1190 ThrGlySerAsnIlePhe...ValLysTyrGluLeu.............. 1200

3612 CGAAGACCGCCGCAACGCCGTTTGGACAAGCGGCATCCGGGACACCAAAC 3661
1201 ................................IleArgAspSer.... 1204

3662 ACTACCGTTCGCAAGATTTCCGCGCCTACCGCCAACAAACCGACCTGCGC 3711
1205 ........AsnGlnAspTyr............................. 1208

3712 CAAATCGGTATGCAGAAAAACCTCGGCAGCGGGCGCGTCGGCATCCTGTT 3761
1209 .....................................ValLeuTh 1211

3762 TTCGCACAACCGGACCGAAAACACCTTCGACGACGGCATCGGCAACTCGG 3811
1211 rArgThrAsnAspValLeuAsnValValThrThrAlaValGlyAsnSerA 1228

3812 CACGGCTTGCCCACGGCGCCGTTTTCGGGCAATACGGCATCG......AC 3855
1228 la...IleAlaAsn.AlaProGlyValSerGlnAsnIleSerArgCysLe 1243

3856 AGGTTCTACATCGGCATCAGCGCGGGCGCGGGTTTTAGCAGCG....... 3898
1243 uGluSerThrAsnThrAlaAlaTyrAsnAsnMetLeuLeuAlaLysAspP 1260

3899 ...........GCAGCCTTTCAGACGGCATCGGAGGCAAAATCCGCCGCC 3937
1260 roSerAspValAlaThrPheValGlyAlaIleAlaThrAspThrSerAla 1276
```

```
    3938 GCGTGCTGCATTACGGCATTCAGGCACGATACCGCGCCGGTTTCGGCGGA 3987
         |||      :::|||:::   :::   :::   :::      :::   ||
    1277 Ala......ValThrThrValAsnLeuAsnAspThrGlnLysThrGlnAs 1291

    3988 TTCGGCATCGAACCGCACATCGGCGCAACGCGCTATTTCGTCCAAAAAGC 4037
         |      |||||||||  :::|||:::   ||||||:::   :::
    1291 pLeuLeuSerAsnArg..LeuGlyThrLeuArgTyrLeuSerAsnAla.. 1306

    4038 GGATTACCGCTACGAAAACGTCAATATCGCCACCCCCGGCCTTGCATTCA 4087
                     |||:::   ::::::|||      |||   |||   :
    1307 .............GluThrSerAspValAla......GlySerAlaThrG 1317

    4088 ACCGCTACCGCGCGGGCATTAAGGCAGATTATTCATTCAAACCGGCGCAA 4137
         ::         :::|||   :::|||:::   |||:::
    1317 lyAlaValSerSerGlyAspGluAlaGluValSerTyr............. 1329

    4138 CACATTTCCATCACGCCTTATTTGAGCCTGTCCTATACCGATGCCGCTTC 4187
            :::         |||:::   :::::::::      |||      ::
    1330 GlyValTrpAlaLysProPheTyrAsnIleAlaGluGlnAspLysLysGl 1346

    4188 GGGC.............AAAGTCCGAACACGC................... 4206
         :|||           |||::::::|||
    1346 yGlyIleAlaGlyTyrLysAlaLysThrThrGlyValValValGlyLeuA 1363

    4207 .............................................GTC 4209
                                                       :::
    1363 spThrLeuAlaSerAspAsnLeuMetIleGlyAlaAlaIleGlyIleThr 1379

    4210 AATACCGCCGTATTGGCTCAGGATTTCGGCAAAACCCGCAGTGCGGAATG 4259
         :::|||   :::      ||||||:::   |||
    1380 LysThrAspIleLysHisGlnAspTyrLysLys.................. 1390

    4260 GGGCGTAAACGCCGAAATCAAAGGTTTCACGCTGTCCCTCCACGCTGCC 4308
          |||   ::::::::::|||:::|||:::::::::|||||:::::::::
    1391 .GlyAspLysThrAspIleAsnGlyLeuSerPheSerLeuTyrGlySer 1406

seq_name: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-268-347-36

seq_documentation_block:
; Sequence 36, Application US/09268347
; Patent No. 6335182
; GENERAL INFORMATION:
;  APPLICANT: Loosmore, Sheena M.
;  TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS
;  FILE REFERENCE: 1038-860
;  CURRENT APPLICATION NUMBER: US/09/268,347
;  CURRENT FILING DATE: 1999-03-16
;  NUMBER OF SEQ ID NOS: 54
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 36
;   LENGTH: 2411
;   TYPE: PRT
;   ORGANISM: Haemophilus influenzae
US-09-268-347-36


alignment_scores:
           Quality:  240.50            Length:   1553
             Ratio:  0.357               Gaps:     80
 Percent Similarity:  43.400   Percent Identity: 19.446

alignment_block:
 US-09-303-518D-649 x US-09-268-347-36   ..

 Align seg 1/1  to: US-09-268-347-36  from: 1  to: 2411

     202 GCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTC........ 243
         :::||||||   ||||||      |||   ||||||   |||
     444 SerLysAspGlnGluVal......LysAlaGlyGluThrValThrPheLy 458

     244 ....GGCAAATCAATG...ACAAAAGCCCCGATGATTGATTTTTCTGTGG 286
             ||||||::::::   :::|||            :::|||:::
```

```
    458 sAlaGlyLysAsnLeuLysValLysGlnAspGlyAlaAsnPheThrTyrS 475

    287 TGTCGCGTAACGGCGTGGCGGCATTGGTGGGC..................... 318
             :::::::::::::::::|||::::::
    475 erLeuGlnAspAlaLeuThrGlyLeuThrSerIleThrLeuGlyGlyThr 491

    319 .....................................GATCAATATAT 329
                                                |||
    492 ThrAsnGlyGlyAsnAspAlaLysThrValIleAsnLysAspGlyLeuTh 508

    330 TGTGAGCGTGGCACATAACGGCGGCTATAACAACGTTGATTTTGGTGCGG 379
         ::::::   |||   ||||||                      |||:::
    508 rIleThrProAlaGlyAsnGly.....................GlyThrT 518

    380 AAGGAAGAAATCCCGATCAACATCGTTTTACTTATAAAATTGTGAAACGG 429
          |||   |||                  |||   ::::::::::|||
    518 hrGlyThrAsn..................ThrIleSerValThrLys... 527

    430 AATAATTATAAAGCAGGGACTAAAGGCCATCCTTATGGCGGCGATTATCA 479
        ::::::   |||||||||:::|||:::
    528 AspGlyIleLysAlaGlyAsnLysAla...................... 536

    480 TATGCCGCGTTTGCATAAATTTGTCACAGATGCAGAACCTGTTGAAATGA 529
                                                  .  :::|
    537 .............................................IleT 538

    530 CCAGTTATATGGATGGGCGGAAATATATCGATCAAAATAATTACCCTGAC 579
        ||:::      :::|||   :::      |||:::   |||:::
    538 hrAsnValAlaSerGlyLeuArgAlaTyrAspAspAlaAsnPhe..... 552

    580 CGTGTTCGTATTGGGGCAGGCAGGCAATATTGGCGATCTGATGAAGATGA 629
                                                    |||
    553 .............................................AspVa 554

    630 GCCCAATAAC......CGCGAAAGTTCATATCATATTGCAAGTGCGTATT 673
            ||||||         :::   :::   |||:::   :::|||||||:
    554 lLeuAsnAsnSerAlaThrAspLeuAsnArgHisValGluAspAlaTyrL 571

    674 CTTGGCTCGTTGGTGGCAATACCTTTGCACAAAATGGATCAGGTGGTGGC 723
        ::                                              |||
    571 ys..........................................Gly 572

    724 ACAGTCAACTTAGGTAGTGAAAAAATTAAACATAGCCCATATGGTTTTTT 773
           :::||||||:::::::::::::   :::   :::|||
    573 LeuLeuAsnLeuAsnGluLysAsnAlaAsnLysGlnProLeuValThrAs 589

    774 ACCAACAGGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATCTATG 823
           |||:::::::::   |||                          |
    589 pSerThrAlaAlaThrValGly.........................A 597

    824 ATGCCCAAAAGCAAAAGTGGTTAATT.........AATGGGGTATTGCAA 864
        ||   :::|||      |||::::::         ||||||:::   :::
    597 spLeuArgLysLeuGlyTrpValValSerThrLysAsnGlyThrLysGlu 613

    865 ACGGGCAACCCCTATATAGGAAAAAGCAATGGCTTCCAGCTGGTTCGTAA 914
           :::|||                                |||:::   :::
    614 GluSerAsn.........................GlnValLysGlnAl 621

    915 AGATTGGTTCTATGATGAAATCTTTGCTGGAGATACCCATTCAGTATTCT 964
         |||         :::   :::|||:::|||          ::::::
    621 aAsp.........GluValLeuPheThrGlyAlaGlyAlaAlaThrValT 635

    965 ACGAACCACGTCAAAATGGGAAATACTCTTTT..................... 996
          :::      :::|||||||||:::::::::
    635 hrSerLysSerGluAsnGlyLysHisThrIleThrValSerValAlaGlu 651

    996 ..................................................... 996

    652 ThrLysAlaAspSerGlyLeuGluLysAspGlyAspThrIleLysLeuLy 668

    997 .............AACGACGAT.................AATAATGGCA 1015
                     |||   |||                |||||||||
    668 sValAspAsnGlnAsnThrAspAsnValLeuThrValGlyAsnAsnGlyT 685
```

```
1016 CA.....  1017
     ||
 685 hrAlaValThrLysGlyGlyPheGluThrValLysThrGlyAlaThrAsp  701

1018 .........GGAAAAATCAATGCCAAACATGAACAC.............  1044
             ||||||::::::::::|||
 702 AlaAspArgGlyLysValThrValLysAspAlaThrAlaAsnAspAlaAs  718

1045 ....................................AATTCTCTGC  1054
                                         ||||||
 718 pLysLysValAlaThrValLysAspValAlaThrAlaIleAsnSerAlaA  735

1055 CTAATAGATTAAAAACACGAACCGTTCAATTGTTTAATGTTTCTTTATCC  1104
       :::   :::||||||::               |||:::::   |||
 735 laThrPheValLysThrGlu.............AsnLeuThrThrSer  746

1105 GAGACAGCAAGAGAACCTGTTTATCATGCT.................GC  1136
                  :::|||:::   :::::                ||
 747 IleAspGluAspAsnProThrAspAsnGlyLysAspAspAlaLeuLysAl  763

1137 AGGTGGTGTCAACAGTTATCGACCCAGACTGAAT.............A  1171
     ||||   :::   :::::::::          |||           :
 763 aGlyAspThrLeuThrPheLysAlaGlyLysAsnLeuLysValLysArgA  780

1172 ATGGAGAAAATATTTCCTTTATTGACGAAGGAAAAGGCGAATTGATACTT  1221
     ::|||:::||||||::|||   |||   :::|||:::      :::
 780 spGlyLysAsnIleThrPhe...AspLeuAlaLysAsnLeuGluValLys  795

1222 ACCAGCAACATCAATCAA......GGTGCTGGAGGATTATATTTCCAAGG  1265
     |||:::::::::::::::            ||||||              ||
 796 ThrAlaLysValSerAspThrLeuThrIleGlyGlyAsnThrProThrGl  812

1266 AGATTTTACGGTCTCGCCTGAAAATAACGAAACTTGGCAAGGCGCGGGCG  1315
     |      |||::::::|||:::   |||   |||      :::   |||:
 812 yGlyThrThrAlaThrProLysValAsnIleThrSerThrAlaAspGlyL  829

1316 TT.....  1317
     ::
 829 euAsnPheAlaLysGluThrAlaAspAlaSerGlySerLysAsnValTyr  845

1317 .....  1317

 846 LeuLysGlyIleAlaThrThrLeuThrGluProSerAlaGlyAlaLysSe  862

1318 ....CATATCAGTGAAGACAGTACCGTTACTTGGAAAGTAAACGGCGTGG  1363
         |||::::::   :::      :::|||   |||   |||::::::::
 862 rSerHisValAspLeuAsnValAspAlaThrLysLysSerAsnAlaAlaS  879

1364 CAAACGACCGCCTGTCCAAAATCGGCAAAGGCACGCTGCACGTTCAAGCC  1413
     ::   :::   :::   :::   |||             :::::|||:::
 879 erIleGluAspValLeuArgAlaGly.........TrpAsnIleGlnGly  892

1414 AAAGGGGAAAACCAAGGCTCGATC.........................  1437
     :::|||:::|||          :::
 893 AsnGlyAsnAsnValAspTyrValAlaThrTyrAspThrValAsnPheTh  909

1438 .......AGCGTGGGCGACGGTACAGTCATTTTGGATCAGCAGGCAGACG  1480
           |||:::|||      ||||||   :::   |||:::||||||
 909 rAspAspSerThrGlyThrThrThrValThrValThrGlnLysAlaAspG  926

1481 ATAAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCTTGGTCAGCGGCAGG  1530
       ||||||
 926 lyLysGly.....  928

1531 GGTACGGTGCAACTGAATGCCGATAATCAGTTCAACCCCGAC........  1572
     :::   |||:::::::::|||   ::::::         |||
 929 AlaAspValLysIleGlyAlaLysThrSerValIleLysAspHisAsnGl  945

1573 .AAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTA..........  1611
      |||||||::   |||                ||||||
 945 yLysLeuPheThrGly..............LysAspLeuLysAspAlaA  957
```

```
1612 ..AACGGGCATTCGCTTTCGTTCCACCGTATTCAAAATACCGATGAAGGG  1659
       ||||||   ::::::|||           ::::::|||      |||
 957 snAsnGlyAlaThrValSerGluAspAspGlyLysAspThrGlyThrGly  973

1660 GCGATGATTGTCAACCACAATCAAGAC......AAAGAATCCACCGTTAC  1703
       :::   :::::         |||      ::::::|||
 974 LeuValThrAlaLysThrValIleAspAlaValAsnLysSerGlyTrpAr  990

1704 CATTACAGGCAATAAAGATATTGCTACAACC...................  1734
      :::||||||:::         |||   |||
 990 gValThrGlyGluGlyAlaThrAlaGluThrGlyAlaThrAlaValAsnA 1007

1735 ..GGCAATAACAACAGCTTGGATAGCAAAAAAGAAATTGCCTAC...AAC  1779
       ||||||   :::::::::   |||      ::::::   :::   |||
1007 laGlyAsnAlaGluThrValThrSerGlyThrSerValAsnPheLysAsn 1023

1780 GGTTGGTTTGGCGAGAAAGATACGACCAAAACGAACGGGCGGCTCAACCT  1829
     |||                  :::::::|||   ||||||::::::|||::
1024 GlyAsnAlaThrThrAlaThrValSerLysAspAsnGlyAsnIleAsnVa 1040

1830 TGTTTAC.....................................C  1837
     :   |||                                     :
1040 lLysTyrAspValAsnValGlyAspGlyLeuLysIleGlyAspAspLysL 1057

1838 AGCCCGCCGCAGAAGACCGCACCCTGCTGCTTTCCGGCGGA.........  1878
     ::   :::|||:::      ||||||   ::::::||||||
1057 ysIleValAlaAspThrThrThrLeuThrValThrGlyGlyLysValSer 1073

1879 ACAAATTTAAACGGCAACATCACGCAAACAAACGGCAAACTGTTTTTCAG  1928
     :::      ::::::|||   :::::::::|||   ||||||      ::
1074 ValProAlaGlyAlaAsnSerValAsnAsnAsnLysLysLeu...ValAs 1089

1929 CGGCAGACCAACACCGCACGCCTACAATCATTTAAACGACCATTGGTCGC  1978
     :::::::              |||   |||:::|||:::      |||:::
1089 nAlaGluGlyLeuAlaThrAlaLeuAsnAsnLeuSer......TrpThrA 1104

1979 AAAAA..............GAGGGCATTCCTCGCGGGGAAATC......  2007
       |||              :::|||      :::||||||
1104 laLysAlaAspLysTyrAlaAspGlyGluSerGluGlyGluThrAspGln 1120

2008 ...GTGTGGGACAACGACTGGATCAACCGCACATTTAAAGCG...GAAAA  2051
       |||      :::|||   :::      ||||||||||||   :::||
1121 GluValLysAlaGlyAspLysVal......ThrPheLysAlaGlyLysAs 1135

2052 CTTCCAAATTAAAGGCGGACAGGCGGTGGTTTCCCGCAATGTTGCCAAAG  2101
     |:::::::::|||   ::::::
1135 nLeuLysValLysGlnSerGlu.........................  1142

2102 TGAAAGGCGATTGGCATTTGAGCAATCACGCCCAAGCAGTTTTTGGTGTC  2151

1142 .....  1142

2152 GCACCGCATCAAAGCCACACAATCTGTACACGTTCGGACTGGACGGGTCT  2201
           ::::::   |||         ::::::      ||||||||
1143 .........LysAspPheThrTyrSerLeuGlnAspThrLeuThrGlyLe 1156

2202 GACAAATTGTGTCGAAAAAACCATTACCGACGATAAAGTGATTGCTTCAT  2251
     ||||               :::|||||||         :::   ::::::
1156 uThr..............SerIleThr.........LeuGlyGlyThrA 1165

2252 TGACTAAGACCGACATCAGCGGCAATGTCGATCTTGCCGATCACGCTCAT  2301
       :::       :::   :::|||:::|||          |||
1165 laAsnGlyArgAsnAspThrGlyThrValIleAsnLysAspGlyLeuThr 1181

2302 TTAAATCTCACAGGGCTTGCCACACTCAACGGCAATCTTAGTGCAAATGG  2351
     ::::::|||::::::   |||:::      |||:::   ::::::|||||
1182 IleThrLeuAlaAsnGlyAlaAlaAla...GlyThrAspAlaSerAsnGl 1197

2352 CGATACACGTTATACAGTCAGCCACAACGCCACCCAAAACGGCAAC....  2397
     |:::|||      :::|||:::   ::::::   :::   ||||||
1197 yAsnThr...IleSerValThrLysAspGlyIleSerAlaGlyAsnLysG 1213

2398 ..CTTAGCCTCGTGGGCAATGCCCAAGCAACATTTAATCAAGCCACATTA 2445
```



```
    3922 GGCAAAATC 3930
         |||||| 
    1749 AspLysIle 1751

seq_name: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-268-347-48

seq_documentation_block:
; Sequence 48, Application US/09268347
; Patent No. 6335182
; GENERAL INFORMATION:
;  APPLICANT: Loosmore, Sheena M.
;  TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS
;  FILE REFERENCE: 1038-860
;  CURRENT APPLICATION NUMBER: US/09/268,347
;  CURRENT FILING DATE: 1999-03-16
;  NUMBER OF SEQ ID NOS: 54
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 48
;   LENGTH: 2048
;   TYPE: PRT
;   ORGANISM: Haemophilus influenzae
US-09-268-347-48


alignment_scores:
            Quality:  234.00            Length:   1480
              Ratio:  0.352               Gaps:     69
 Percent Similarity:  44.865   Percent Identity: 20.338

alignment_block:
 US-09-303-518D-649 x US-09-268-347-48   ..

 Align seg 1/1  to: US-09-268-347-48  from: 1  to: 2048

     163 CGCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTC.............GG 200
         :::|||   ::::::|||   |||      ::::::           ||
     657 LysAspThrThrLysAsnAlaGlyAlaValSerIleLeuLysLeuLysGl 673

     201 GGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTC......G 244
         |   :::   :::   |||   :::||||||         |||      |
     673 yLysAsnGlyLeuThrValAlaThrLysLysAspGlyThrValThrPheG 690

     245 GCAAATCAATGACAAAAGCCCCGATGATTGATTTTTCTGTGGTGTCGCGT 294
         ||  |||:::   ::::::      |||      |||:::::::::::::
     690 lyLeuSerGlnAspSerGlyLeuThrIleGlyLysSerThrLeuAsnAsn 706

     295 AACGGCGTGGCGGCATTGGTGGGCGATCAATATATTGTGAGCGTGGCACA 344
         :::|||:::::::::         ::::::   |||         |||::
     707 AspGlyLeuThrValLysAspThrAsnGluGlnIleGlnValGlyAlaAs 723

     345 TAACGGCGGCTATAACAACGTTGATTTTGGTGCGGAAGGAAGAAATCCCG 394
         ::::      ::::::||||||:::           |||   ||||||
     723 nGlyIleLysPheThrAsnValAsn.............GlySerAsnProG 736

     395 ATCAACATCGTTTTACTTATAAAATTGTGAAACGGAATAATTATAAAGCA 444
                   :::         :::   |||:::      |||
     736 lyThrGlyIleAlaAsnThrAlaArgIleThrArgAsp......LysIle 750

     445 GGGACTAAAGGCCATCCTTATGGCGGCGATTATCATATGCCGCGTTTGCA 494
         |||      |||         :::   |||   :::   |||
     751 GlyPheAlaGlySerAspGlyAlaValAspThrAsnLysPro........ 764

     495 TAAATTTGTCACAGATGCAGAACCTGTTGAAATGACCAGTTATATGGATG 544

     764 .................................................. 764

     545 GGCGGAAATATATCGATCAAAATAATTACCCTGACCGTGTTCGTATTGGG 594
               |||:::||||||           |||:::::::::::::
     765 ........TyrLeuAspGln.............AspLysLeuGlnVal... 773

     595 GCAGGCAGGCAATATTGGCGATCTGATGAAGATGAGCCCAATAACCGCGA 644

     773 .................................................. 773

     645 AAGTTCATATCATATTGCAAGTGCGTATTCTTGGCTCGTTGGTGGCAATA 694
                                                       ||||||:
     774 ..........................................GlyAsnV 776

     695 CCTTTGCACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGTGAA 744
         ::         |||   :::            :::|||   |||::::::
     776 alLysIleThrAsnThrGly............IleAsnAlaGlyGlyLys 788

     745 AAAATTAAACATAGCCCATATGGTTTTTTACCAACAGGAGGCTCATTTGG 794
            |||                |||:::   ||||||      |||:::::
     789 AlaIleThr............GlyLeuSerProThrLeuProSerIleAl 801

     795 CGACAGTGGCTCACCAATGTTT..............ATCTATGATGCCC 829
         :|||::::::|||      :::               |||   |||   :
     801 aAspGlnSerSerArgAsnIleGluLeuGlyAsnThrIleGlnAspLysA 818

     830 AAAAGCAAAAGTGGTTA...ATTAATGGGGTATTGCAAACGGGCAACCCC 876
         ::|||::::::         ||||||   :::|||:::||||||
     818 spLysSerAsnAlaAlaSerIleAsnAspIleLeuAsnThrGlyPheAsn 834

     877 TATATAGGAAAAAGCAATGGCTTCCAGCTGGTTCGTAAAGATTGGTTCTA 926
               :::::::::|||   :::::::::|||                ||
     835 LeuLysAsnAsnAsnAsnProIleAspPheValSerThr.........Ty 848

     927 TGATGAAATC........................................ 936
         ||||   :::
     848 rAspIleValAspPheAlaAsnGlyAsnAlaThrThrAlaThrValThrH 865

     937 .....TTTGCTGGAGATACCCATTCAGTATTCTACGAACCACGTCAAAAT 981
              |||:::   |||   :::|||   |||:::   :::   :::
     865 isAspThrAlaAsnLysThrSerLysValValTyrAspValAsnValAsp 881

     982 GGG.........AAATACTCTTTTAACGACGATAATAATGGCACAGGAAA 1022
                      :::   :::|||||||||:::      |||
     882 AspThrThrIleHisLeuThrGlyThrAspAspAsnLysLysLeuGlyVa 898

    1023 AATCAATGCCAAACATGAACACAATTCTCTGCCTAATAGATTAAAAACAC 1072
         ::::::|||   :::   :::|||      ::::::      |||:
     898 lLysThrThrLysLeuAsnLysThrSerAlaAsnGlyAsnThrAlaThrA 915

    1073 GAACCGTTCAATTGTTTAATGTTTCTTTATCCGAG.............. 1107
         ::           |||||||||:::   |||:::
     915 sn...........PheAsnValAsnSerSerAspGluAspAlaLeuVal 927

    1108 ..........................ACAGCAAGAGAACCTGTTTA 1127
                                    |||      :::   :::::
     928 AsnAlaLysAspIleAlaGluAsnLeuAsnThrLeuAlaLysGluIleHi 944

    1128 TCATGCTGCAGGTGGTGTCAACAGTTATCGACCCAGACTG.......... 1167
         :   :::   |||   :::::::::          :::
     944 sThrThrLysGlyThrAlaAspThrAlaLeuGlnThrPheThrValLysL 961

    1168 ..............AATAATGGAGAAAATATTTCCTTTATTGACGAAGGA 1203
                        ||||||:::::::::   :::   |||      |||
     961 ysValAspGluAsnAsnAsnAlaAspAspAlaAsnAlaIleThrValGly 977

    1204 AAAGGCGAATTGATACTTACCAGCAACATCAATCAAGGTGCTGGAGGATT 1253
         :::                    :::|||   |||:::   ::::::   ||
     978 Gln..................LysAsnAlaAsnAsnGlnValAsnThrLe 988

    1254 ATATTTCCAAGGAGATTTTACGGTCTCGCCTGAAAATAACGAAACTTGGC 1303
         |   ::::::|||:::                                 :
     988 uThrLeuLysGlyGlu................................A 994

    1304 AAGGCGCGGGCGTTCATATCAGTGAAGACAGTACCGTTACTTGGAAAGTA 1353
         ::|||   ::::::      :::::::::::|||||||||:::   :::
     994 snGlyLeuAsnIleLysThrAspLysAsnGlyThrValThrPheGlyIle 1010

    1354 AACGGCGTGGCAAACGACCGCCTGTCCAAAATCGGCAAAGGCACGCTGCA 1403
         |||   :::::::::            |||   ||||||:::||||||
    1011 AsnThrThrSerGly.........LeuLysAlaGlyLysSerThrLeu.. 1023
```



```
1573 pSerSerGlyGlnAlaLysAlaAsn.....ThrProValLeuS 1586

2852 CGGTAGAATCCCGTTTCAACACGCTGACGGTAAACGGCAAATTG..... 2895
1586 erAla.....AsnGlyLeuAspLeuGlyGlyLysValIleSer 1598

2896 ...AACGGTCAGGGAACATTCCGCTTTATGTCGGAACTCTTCGGCTACCG 2942
1599 AsnValGlyLysGlyThr.....LysAs 1606

2943 CAGCGACAAATTGAAGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGG 2992
1606 pThrAspAlaAlaAsnValGlnGlnLeuAsnGluValArgAsnLeuLeuG 1623

2993 CGGTCAACAATACCGGCAACGAACCTGCAAGCCTCGAACAATTGACGGTA 3042
1623 lyLeuGlyAsnAlaGlyAsnAspAsnAlaAspGlyAsnGlnValAsnIle 1639

3043 GTGGAAGGAAAAGACAACAAACCGCTGTCCGAAAACCTTAATTTCACCCT 3092
1640 AlaAspIleLysLysAspProAsnSerGlySerSerSerAsnArgThrVa 1656

3093 G.....CAAAACGAACACGTCG 3109
1656 lIleLysAlaGlyThrValLeuGlyGlyLysGlyAsnAsnAspThrGluL 1673

3110 ATGCCGGCGCGTGGCGTTACCAACTC...ATCCGCAAAGACGGCGAGTTC 3156
1673 ysLeuAlaThrGlyGlyIleGlnValGlyValAspLysAspGlyAsnAla 1689

3157 CGC......CTGCATAATCCG...GTCAAAGAACAAGAGCTTTCCGACAA 3197
1690 AsnGlyAspLeuSerAsnValTrpValLysThrGlnLysAspGlySerLy 1706

3198 ACTCGGCAAGGCAGAAGCCAAAAAACAGGCGGAAAAAGACAAC..... 3240
1706 sLysAlaLeuLeuAlaThrTyrAsnAlaAlaGlyGlnThrAsnTyrLeuT 1723

3241 ..........GCGCAAAGCCTTGACGCGCTGATTGCGGCCGGGCGCGAT 3279
1723 hrAsnAsnProAlaGluAlaIleAspArgIleAsnGluGlnGlyIleArg 1739

3280 GCCGTCGAAAAGACAGAAAGCGTTGCCGAACCGGCCCGGCAGGCAGGCGG 3329
1740 PhePheHisValAsnAspGlyAsnGlnGluProValValGln......Gl 1754

3330 GGAAAATGTCGGCATTATGCAGGCGGAGGAAGAGAAAAAACGGGTG.... 3375
1754 yArgAsn...GlyIleAspSerSerAlaSerGlyLysHisSerValAlaI 1770

3376 .............CAGGCGGATAAAGACACCGCCTTGGCGAAACAGCGC 3411
1770 leGlyPheGlnAlaLysAlaAspGlyGluAlaAlaValAlaIleGlyArg 1786

3412 GAAGCGGAAACCCGGCCGGCTACCACCGCCTTCCCCCGCGCCCGCCGCGC 3461
1787 GlnThrGlnAlaGlyAsnGlnSerIleAlaIleGlyAspAsnAlaGlnAl 1803

3462 CCGCCGGGATTTGCCGCAACTGCAACCCCAACCGCAGCCCCAACCGCAGC 3511
1803 aThrGlyAsp.....GlnS 1808

3512 GCGACCTGATCAGCCGTTAT.....GCCAATAGCGGTTTG 3546
1808 erIleAlaIleGlyArgThrAsnValValAlaGlyLysHisSerGlyAla 1824

3547 AGTGAATTTTCCGCCACGCTCAACAGCGTTTTCGCCGTACAGGACG...A 3593
1825 .IleGlyAspProSerThrValLysAlaAspAsnSerTyrSerValGlyA 1841

3594 ATTAGACCGCGTATTTGCCGAAGACCGCCGCAACGCCGTTTGGACAAGCG 3643
1841 snAsnAsnGlnPheThrAspAlaThrGlnThrAspValPheGlyValGly 1857

3644 GCATCCGGGACACCAAACACTACCGTTCGCAAGATTTCCGCGCCTACCGC 3693
1858 .....AsnAsnIleThrValThrGluSerAsnSerValAlaLe 1870

3694 CAACAAACCGACCTGCGCCAAATCGGTATGCAGAAAAACCTCGGCAGCGG 3743
1870 uGly.....SerAsnSerAlaIleSerAlaGlyThrHisAlaG 1883

3744 GCGCGTCGGCATCCTGTTTTCGCACAACCGGACCGAAAACACCTTCGACG 3793
1883 lyThrGlnAlaLys.....LysSerAspGlyThrAlaGlyThrThrThr 1897

3794 ACGGCATCGGCAACTCGGCACGGCTTGCCCACGGCGCCGTTTTCGGGCAA 3843
1898 ThrAlaGlyAlaThrGly.....ThrValLysGlyPheAlaGlyGl 1911

3844 TACGGCA.....TCGACAGGTTCTACATCGGCATCAGCGCGGG 3881
1911 nThrAlaValGlyAlaValSerValGlyAlaSerGlyAlaGluArgArgI 1928

3882 CGCGGGTTTTAGCAGCGGCAGCCTTTCAGACGGCATCG 3919
1928 leGlnAsnValAlaAlaGlyGluValSerAlaThrSer 1940
```

```
seq_name: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-268-347-49

seq_documentation_block:
; Sequence 49, Application US/09268347
; Patent No. 6335182
; GENERAL INFORMATION:
;  APPLICANT: Loosmore, Sheena M.
;  TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS
;  FILE REFERENCE: 1038-860
;  CURRENT APPLICATION NUMBER: US/09/268,347
;  CURRENT FILING DATE: 1999-03-16
;  NUMBER OF SEQ ID NOS: 54
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 49
;   LENGTH: 2314
;   TYPE: PRT
;   ORGANISM: Haemophilus influenzae
US-09-268-347-49


alignment_scores:
            Quality:  232.50              Length:   1750
              Ratio:  0.305                 Gaps:     87
 Percent Similarity:  43.600   Percent Identity: 19.143

alignment_block:
 US-09-303-518D-649 x US-09-268-347-49   ..

 Align seq 1/1  to: US-09-268-347-49  from: 1  to: 2314

      541 GATGGGCGGAAATATATCGATCAAAATAAT......TACCCTGACCGTGT 584
      460 AspGlyLeuLysPheThrAspAsnSerAsnThrAlaLeuGluAspThrTh 476

      585 TCGT.....ATTGGGGCAGGCAGGCAATATTGGCGATCTG 619
      476 rArgIleThrLysAspLysIleGlyPheSerAsnLysAlaGlyThrValA 493

      620 ATGAAGATGAGCCCAATAACCGCGAAAGTTCATATCATATTGCAAGTGCG 669
      493 spGluAsnLysProTyrLeuAspLysAspLysLeuLysValGlyAsn... 508

      670 TATTCTTGGCTCGTTGGTGGCAATACCTTTGCACAAAATGGATCAGGTGG 719
      509 ...SerThrLeuAsnAsnGlyGlyLeuThrValAsnAsnThrIleGlyGl 524

      720 TGGC..... 723
```

```
    |:::
  524 ySerAsnLysGlnIleGlnValGlyAlaAspGlyIleLysPheAlaAspV 541

  724 ..ACAGTCAACTTA.................GGTAGTGAAAAAATTAAA 753
        :::|||||||:::                 |||:::   :::|||
  541 alAsnValAsnValSerAsnAlaAlaLysPheGlyThrThrArgIleThr 557

  754 CATAGCCCATATGGTTTTTTACCAACAGGAGGCTCATTTGGCGACAGTGG 803
      ::::::      ||||||      :::   |||:::      |||:::
  558 GluGluGluIleGlyPheAlaAspAlaAspGlyLysVal...AspLysLy 573

  804 CTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGGTTAATTAATG 853
       ||||||        |||     :::||||||
  573 sSerPro.........TyrLeuAspLysLysGln.............. 581

  854 GGGTATTGCAAACGGGCAACCCCTATATAGGAAAAAGCAATGGCTTCCAG 903
        ||||||:::|||:::      |||   |||:::::::|||:::::
  582 .....LeuGlnValGlyGlyValLysIleThrLysAspSerGlyIleAsn 596

  904 CTGGTTCGTAAAGATTGGTTCTATGATGAAATCTTTGCTGGAGATACCCA 953
            :::                :::      :::   ||||||
  597 AlaGlyAspGlnLysIleSerAsnValLysAspAlaThrAspAspThrAs 613

  954 TTCAGTATTCTACGAACCACGTCAAAAT..................... 981
       :::|||   ||:|:::      ::::::
  613 pAlaValThrTyrLysGlnLeuLysGlnValGlnGlnAspAlaAspGlyA 630

  982 .....GGGAAATACTCTTTTAACGACGATAATAATGGCACAGGAAAAATC 1026
            ::::::|||:::::::|||:::::::::             |||
  630 laLeuGlnSerPheSerIleArgAspGluLysGlyGlnGluPheThrIle 646

 1027 AATGCCAAACATGAACACAATTCTCTGCCTAATAGATTAAAAACACGAAC 1076
      :::      :::::::::::::::   ||||||   ::::::|||   ||
  647 SerAsnLeuTyrSerAsnGlyAsnThrProAsnThrPheGluThrIleTh 663

 1077 CGTTCAATTGTTTAAT...GTTTCTTTATCC...GAGACAGCAAGAGAAC 1120
      |            |||   :::|||:::|||   :::   |||:::
  663 rPheAlaGlyGluAsnGlyIleSerIleSerAsnAspIleAlaLysGlyL 680

 1121 CTGTTTATCATGCTGCAGGTGGTGTCAACAGTTAT...CGACCCAGACTG 1167
        |||      :::        :::|||:::        |||:::|||
  680 ysValLysValGlyIleAspProIleAsnGlyLeuThrThrProLysLeu 696

 1168 AATAATGGAGAAAATATTTCCTTTATTGACGAAGGAAAAGGCGAATTGAT 1217
      :::   |||::::::              :::||||||  :::|||::
  697 ThrValGlySerAsp............LysAspGlyLysThrGlnLeuVa 709

 1218 ACTT........................ACCAGCAACATCAATCAAGGT. 1242
      ::::                        |||:::||||||   :::|||
  709 lIleGluGlnValAlaSerGlyAsnGlyThrLysAsnIleIleArgGlyV 726

 1243 ............................GCTGGAGGATTATATTTC... 1260
                                  |||||||||:::
  726 alSerProThrLeuProSerIleThrAsnAlaGlyGlyValArgThrThr 742

 1261 ...CAAGGAGATTTTACGGTCTCGCCTGAA.................... 1287
         ||||||:::      :::|||   |||
  743 GluGlnGlyAsnThrIleThrSerAspGluAspLysSerLysAlaAlaSe 759

 1288 ....................................AATAACGAAACTT 1300
                                          ||||||::::::
  759 rIleGlyAspIleLeuAsnThrGlyPheAsnLeuLysAsnAsnSerAsnS 776

 1301 GGCAAGGCGCGGGCGTTCATATCAGTGAAGACAGTACCGTTACTTGGAAA 1350
            |||           :::|||      :::||||||   :::
  776 erValGly.........PheValSerThrTyrAsnThrValAspPhe... 788

 1351 GTAAACGGC..................GTGGCAAACGACCGCCTGTCCAA 1382
      :::::::|||                 |||:::   |||:::   :::::
  789 IleAspGlyAsnAlaThrThrAlaLysValThrTyrAspGluThrAsnGl 805

 1383 AATCGGCAAAGGCACGCTGCACGTT.....................CAAG 1411
      :   :::|||   |||      |||                    :::
```

```
  805 nThrSerLysValThrTyrAspValAsnValAspGluLysThrIleGluL 822

 1412 CCAAAGGGGAAAACCAAGGCTCGATCAGCGTGGGCGACGGTACAGTCATT 1461
           |||:::|||      :::   :::::::|||      |||:::
  822 euThrGlyAspAsnGlyLysThrAsnLysIleGlyValLysThrThrThr 838

 1462 TTGGATCAGCAGGCAGACGATAAAGGCAAAAAACAAGCCTTTAGT..... 1506
      |||          ::::::   :::|||||||          ||||||
  839 Leu...ThrThrThrAsnAlaAsnGlyLysAlaThrAsnPheSerThrTh 854

 1506 ..................................................... 1506

  854 rAspAsnAspAlaLeuValAsnAlaLysAspIleAlaGluAsnLeuAsnT 871

 1507 ..........GAAATCGGCTTGGTCAGCGGCAGGGGT............ 1533
                 ||||||      :::::::|||   :::
  871 hrLeuAlaLysGluIleHisThrThrLysGlyThrAlaAspThrAlaLeu 887

 1534 ...ACGGTGCAACTGAATGCCGATAATCAGTTCAACCCCGACAAACTCTA 1580
         |||   :::::::::   |||:::      :::   :::   :::
  888 GlnThrPheLysValLysLysAspGlyAlaThrAspAspGluThrIleTh 904

 1581 TTTCGGCTTTCGCGGCGGACGTTTGGATTTAAACGGGCATTCGCTTTCGT 1630
         |||      |||   :::         ||||||   :::::::::
  904 rValGlyLysAspGlyThrGln.........AsnGlyLysThrValAsnT 918

 1631 TCCACCGTATTCAAAATACCGATGAAGGGGCGATGATTGTCAACCACAAT 1680
            ::::::::::      |||:::::::::   |||      |||
  918 hrLeuLysLeuLysGly......GluAsnGlyLeuThrValAlaThrAsn 932

 1681 CAAGACAAAGAATCCACCGTTACCATT..................... 1707
           |||::::::||||||||||:::
  933 ......LysAspGlyThrValThrPheGlyIleAsnThrGlnSerGlyLe 947

 1708 ...................ACAGGCAATAAAGAT.........ATTGCTA 1729
                          |||   |||||||||         |||   :
  947 uLysAlaGlyAspSerThrThrLeuAsnLysAspGlyLeuSerIleLysA 964

 1730 CAACCGGCAATAACAACAGCTTGGAT..................... 1755
      ::   ::::::|||:::::::::::
  964 snProAlaSerAsnGluGlnIleGlnValGlyAlaAspGlyValLysPhe 980

 1755 ..................................................... 1755

  981 AlaLysValAspLysGlyAsnSerSerThrGlyIleAspGlyThrSerAr 997

 1756 ....AGCAAAAAAGAAATTGCCTACAACGGTTGGTTTGGCGAGAAAGATA 1801
          :::|||   :::|||:::::::::|||      |||:::   ||||
  997 gIleThrLysAspGlnIleGlyPheThrGlyAlaAsnGlySerLeuAspT 1014

 1802 CGACCAAAACGAAC.....................GGGCGGCTCAAC 1827
      ||||||||   :::                     |||:::::::::
 1014 hrThrLysProHisLeuThrLysAspLysLeuLysValGlyGluValGlu 1030

 1828 CTTGTTTACCAGCCCGCCGCAGAAGACCGCACCCTGCTGCTTTCCGGCGG 1877
      ::::::                      :::|||   :::   :::||||
 1031 IleThr.....................AsnThrGlyIleAsnAlaGlyGl 1040

 1878 A.........ACAAATTTA...AACGGCAACATCACGCAAACAAACGGCA 1915
      |        ||||||:::   :::|||:::||||||||:::::::::
 1040 yLysLysIleThrAsnIleGlnSerGlyAspIleThrGlnAsnSerAsnA 1057

 1916 AACTGTTTTTCAGCGGCAGA............................ 1935
                ::::||||||
 1057 spAlaValThrGlyGlyArgValTyrAspLeuLysThrGluLeuGluSer 1073

 1936 ............CCAACACCGCACGCCTACAATCATTTAAACGAC..... 1968
                   ::      |||   |||   |||::::::
 1074 LysIleAsnSerAlaAlaLysThrAlaGlnAsnSerLeuHisGluPheSe 1090

 1969 .....................CATTGGTCGCAAAAAGAGGGCATTCCTC 1996
                           |||::::::   ::::::
 1090 rValAlaAspGluGlnGlyAsnHisPheThrValSerAsnProTyrSerS 1107
```


1997 GCGGGGAAATCGTGTGGGACAACGACTGGATCAACCGCACATTTAAAGCG 2046
1107 erTyrAspThrSerLysThrSerAspValIle.....ThrPheAlaGly 1121
2047 GAAAAC...TTCCAAATTAAAGGCGGACAGGCGGTGGTTTCCCGCAATGT 2093
1122 GluAsnGlyIleThrThrLysValAsnLysGlyValValArgValGlyIl 1138
2094 TGCCAAAGTGAAAGGCGATTGGCATTTG..... 2121
1138 eAspGlnThrLysGly.....LeuThrThrProLysLeuThrValG 1152
2122 ..AGCAATCACGCCCAAGCAGTTTTTGGTGTCGCACCGCATCAAAGCCAC 2169
1152 lyAsnAsnAsnGlyLysGlyIleValIleAspSerLysAspGlyGlnAsn 1168
2170 ACAATCTGTACACGTTCGGACTGGACGGGTCTGACAAATTGTGTCGAAAA 2219
1169 ThrIle.....ThrGlyLeuSerAsnThrLeu...Al 1178
2220 AACCATTACCGACGATAAAGTGATTGCTTCATTGACTAAG..... 2259
1178 aAsnValThrAsnAspGlyAlaGlyHisAlaLeuSerGlnGlyLeuAlaA 1195
2260 ACCGACATCAGCGGCAATGTCGATCTTGCCGAT.....CACGCT 2298
1195 snAspThrAspLysThrArgAlaAlaSerIleGlyAspValLeuAsnAla 1211
2299 CATTTAAATCTCACAGGGCTTGCCACACTCAACGGC.....AATCT 2339
1212 GlyPheAsnLeuGlnGly.....AsnGlyGluAlaValAspPh 1224
2340 TAGTGCAAATGGCGATACACGTTATACAGTCAGCCACAACGCCACC.... 2385
1224 eValSerThrTyrAspThrValAspPheIleAspGlyAsnAlaThrThrA 1241
2385 2385
1241 laLysValThrTyrAspAspThrSerLysThrSerLysValValTyrAsp 1257
2386 CAAAACGGCAACCTTAGCCTC.....GTGGG 2411
1258 ValAsnValAspAsnLysThrIleGluValThrSerAspLysLysLeuGl 1274
2412 CAATGCCCAAGCAACATTTAATCAAGCCACATTAAACGGCAAC..... 2454
1274 yValLysThrThrThrLeuThrLysThrSerAlaAsnGlyAsnAlaThrL 1291
2455 ACATCGGCTTCGGGCAATGCTTCATTTAATCTAAGCGACCACGCC 2499
1291 ysPheSerAlaAlaAspGlyAspAlaLeuValLysAlaSerAspIleAla 1307
2500 GTACAAAACGGCAGTCTGACGCTTTCCGGCAAC..... 2532
1308 ThrHisLeuAsn.....ThrLeuSerGlyAspIleGlnThrAlaLysGl 1322
2533 GCTAAGGCAAACGTAAGCCATTCCGCACTCAACGGTAATG 2572
1322 yAlaSerGlnAlaSerSerSerAlaSerTyrValAspAlaAspGlyAsnL 1339
2573 TCTCCCTAGCCGATAAGGCAGTATTCCATTTTGAAAGCAGCCGCTTTACC 2622
1339 ysValIleTyrAspSerThrAspLysLysTyrTyrGlnValAsnAspLys 1355
2623 GGACAAATCAGCGGCGGCAAGGATACGGCATTACACTTAAAAGACAGCGA 2672
1356 GlyGlnValAspLysAsnLysGluValAlaLysAspLysLeuValAlaGl 1372
2673 ATGGACGCTGCCGTCAGGCACGGAATTAGGCAATTTAAACCTTGACAACG 2722
1372 nAlaGlnThrProAspGlyThr...LeuAlaGlnMetAsnValLysServ 1388
2723 CCACCATT.....ACACTCAATTCCGCCTATCGCCACGATGCGGCAGGG 2766
1388 alIleValLysGluGlnValAsnAspAlaAsnLysLysGlnGlyIleAsn 1404
2767 GCGCAAACC.....GGCAGTGCGACAGATGCGCC 2795
1405 GluAspAsnAlaPheIleLysGlyLeuGluAsnAlaAlaLysAspThrLy 1421
2796 GCGCCGCCGTTCGCGCCGTTCGCGCCGTTCCCTATTATCCGTT.....A 2839
1421 sThrLysAsnAlaAlaValThrValGlyAspLeuAsnAlaValAlaGlnT 1438
2840 CACCG.....CCAACTTCGGTAGAATCCCGTTTC 2868
1438 hrProLeuThrPheAlaGlyAspThrGlyThrThrAlaLysLysLeuGly 1454
2869 AACACGCTGACGGTAAACGGC..... 2889
1455 GluThrLeuThrIleLysGlyGlyGlnThrAspThrAsnLysLeuThrAs 1471
2889 2889
1471 pAsnAsnIleGlyValValAlaGlyThrAspGlyPheThrValLysLeuA 1488
2890 AAATTGAACGGTCAGGGAACA..... 2910
1488 laLysAspLeuThrAsnLeuAsnSerValAsnAlaGlyGlyThrArgIle 1504
2911 TTCCGCTTTATGTCG..... 2925
1505 AspGluLysGlyIleSerPheValAspAlaAsnGlyGlnAlaLysAlaAs 1521
2926 GAACTCTTCGGCTACCGC.... 2943
1521 nThrProValLeuSerAlaAsnGlyLeuAspLeuGlyGlyLysArgIleS 1538
2944 AGCGACAAATTGAAGCTGGCGGAA 2967
1538 erAsnIleGlyAlaAlaValAspAspAsnAspAlaValAsnPheLysGln 1554
2968 AGTTCCGAAGGCACTTACACCTTG...GCGGTCAACAATACCGGCAACGA 3014
1555 PheAsnGluValAlaLysThrValAsnAsnLeuAsnAsnGlnSerAsnSe 1571
3015 ACCTGCAAGCCTCGAACAATTGACGGTA.....GTGG 3046
1571 rGlyAlaSerLeuProPheValValThrAspAlaAsnGlyLysProIleA 1588
3047 AAGGAAAAGACAACAAACCGCTGTCCGAAAACCTTAATTTCACCCTGCAA 3096
1588 snGlyThrAspGlyLysProGlnLysAlaIleLysGlyAlaAspGlyLys 1604
3097 AACGAACACGTCGATGCCGGCGCGTGGCGTTACCAACTCATCCGCAAAGA 3146
1605 TyrTyrHisAlaAsnAlaAsnGly.....ValProValAspLysAs 1618
3147 CGGCGAGTTCCGCCTGCATAATCCGGTCAAAGAACAAGAGCTTTCCGACA 3196
1618 pGlyLys.....ProIleThrAsp.....AlaAspL 1627
3197 AACTC..... 3201
1627 ysLeuAlaAsnLeuAlaAlaHisGlyLysProLeuAspAlaGlyHisGln 1643
3202 GGCAAGGCAGAAGCC..... 3216
1644 ValValAlaSerLeuGlyGlyAsnSerAspAlaIleThrLeuThrAsnIl 1660
3217 AAAAAACAGGCGGAAA 3232
1660 eLysSerThrLeuProGlnIleAspThrProAsnThrGlyAsnAlaAsnA 1677
3233 AAGACAACGCGCAAAGCCTTGACGCGCTGATTGCGGCCGGGCGCGATGCC 3282

```
          :::|||||||||||||   :::|||   ||||||   ::::::
1677 laGlyGlnAlaGlnSerLeuProSerLeuSerAlaAlaGlnGlnSerAsn 1693

3283 GTCGAAAAGACAGAAAGCGTTGCCGAACCG..................... 3312
     :::   ::::::::::::|||   :::
1694 AlaAlaSerValLysAspValLeuAsnValGlyPheAsnLeuGlnThrAs 1710

3313 ..............................................GCCC 3316
                                                    ::::
1710 nHisAsnGlnValAspPheValLysAlaTyrAspThrValAsnPheValA 1727

3317 GGCAGGCAGGCGGGGAAAATGTCGGCATTATGCAGGCGGAGGAAGAGAAA 3366
     ::   :::|||::::::   :::::::::   :::|||:::
1727 snGlyThrGlyAlaAspIleThrSerValArgSerAlaAspGlyThrMet 1743

3367 AAACGGGTGCAGGCG.......................GATAAAGACAC 3392
     :::::::::   :::                      |||      ::
1744 SerAsnIleThrValAsnThrAlaLeuAlaAlaThrAspAspAspGlyAs 1760

3393 CGCCTTGGCGAAACAGCGCGAAGCGGAAACCCGGCCGGCTACCACCGCCT 3442
     ::::|||   |||   ::::::::::::         |||         :
1760 nValLeuIleLysAlaLysAspGlyLysPheTyrLysAlaAspAspLeuM 1777

3443 TCCCC............CGCGCCCGCCGCGCCCGCCGGGATTTGCCGCAA 3480
     ::|||           :::|||   ::::::      |||
1777 etProAsnGlySerLeuLysAlaGlyLysSerAlaSerAsp......... 1790

3481 CTGCAACCCCAACCGCAGCCCCAACCGCAGCGCGACCTGATCAGCCGTTA 3530
                :::                    :::|||::::::
1791 ............AlaLysThrProThrGlyLeuSerLeuValAsnProAs 1803

3531 TGCCAATAGCGGTTTGAGTGAATTTTCCGCCACGCTCAAC.......... 3570
      |||:::::::|||   :::      :::::::::||||||
1803 nAlaGlyLysGlySerThrGlyAspAlaValAlaLeuAsnAsnLeuSerL 1820

3571 ..AGCGTTTTCGCCGTACAGGAC.....................GAATTA 3597
       :::||||||      :::|||                     :::
1820 ysAlaValPheLysSerLysAspGlyThrThrThrThrThrValSerSer 1836

3598 GACCGCGTATTTGCCGAAGACCGCCGCAACGCCGTTTGGACA......... 3639
     |||   :::      :::   :::   |||:::      |||
1837 AspGlyIleSerIleGlnGlyLysAspAsnSerSerIleThrLeuSerLy 1853

3640 ..........................................AGCGGCA 3646
                                                :::|||
1853 sAspGlyLeuAsnValGlyGlyLysValIleSerAsnValGlyLysGlyT 1870

3647 TCCGGGACACCAAACACTACCGTTCGCAAGATTTCCGCGCCTACCGC... 3693
       :::||||||         :::   |||:::::::      |||
1870 hrLysAspThrAspAlaAlaAsnValGlnGlnLeuAsnGluValArgAsn 1886

3694 .........................................CAACAAAC 3701
                                               :::|||::
1887 LeuLeuGlyLeuGlyAsnAlaGlyAsnAspAsnAlaAspGlyAsnGlnVa 1903

3702 CGACCTGCGCCAAATCGGTATGCAGAAAAACCTCGGCAGCGGGCGCGTCG 3751
     :::::::   :::|||      :::   |||   ||||||
1903 lAsnIleAlaAspIleLysLysAspProAsnSerGlySer.......... 1916

3752 GCATCCTGTTTTCGCACAACCGGACC.........GAAAACACCTTCGAC 3792
               |||   |||||||||            :::|||
1917 ...........SerSerAsnArgThrValIleLysAlaGlyThrValLeu 1929

3793 GACGGCATCGGCAAC......TCGGCACGGCTTGCCCACGGCGCCGTTTT 3836
        |||   ||||||      :::   :::||||||   |||:::|||
1930 GlyGlyLysGlyAsnAsnAspThrGluLysLeuAlaThrGlyGlyVal.. 1945

3837 CGGGCAATACGGCATCGACAGGTTCTACATCGGCATCAGCGCGGGCGCGG 3886
         |||   |||:::|||:::                          |
1946 ....GlnValGlyValAspLys.....................AspG 1953

3887 GTTTTAGCAGCGGCAGCCTTTCAGACGGCATCGGAGGCAAAATCCGCCGC 3936
     ||   ::::::|||:::|||||||:::   :::      |||   ::::::
```

```
1953 lyAsnAlaAsnGlyAspLeuSerAsn...ValTrpValLysThrGlnLys 1968

3937 CGCGTGCTGCATTACGGCATTCAGGCACGATACCGCGCCGGTTTCGGCGG 3986
                     ::::::   |||   |||:::|||:::
1969 AspGlySerLysLysAlaLeuLeuAlaThrTyrAsnAlaAla........ 1982

3987 ATTCGGCATCGAACCGCACATCGGCGCAACGCGCTATTTCGTCCAAAAAG 4036
                              |||   |||:::|||::::::::::::
1983 .....................GlyGlnThrAsnTyrLeuThrAsnAsnP 1992

4037 CGGATTACCGCTACGAAAACGTCAATATCGCCACCCCCGGCCTTGCATTC 4086
                     :::::::::|||            |||:::   |||
1992 roAlaGluAlaIleAspArgIleAsn......GluGlnGlyIleArgPhe 2006

4087 AACCGCTAC.............................CGCGCGGG 4103
        :::                                  |||   ||
2007 PheHisValAsnAspGlyAsnGlnGluProValValGlnGlyArgAsnGl 2023

4104 CATTAAGGCAGATTATTCATTCAAACCGGCGCAACACATTTCCATCACGC 4153
     ||||   ::::::   |||   |||   :::
2023 yIleAspSerSerAlaSerGlyLysHisSer.................. 2033

4154 CTTATTTGAGCCTGTCCTATACCGATGCCGCTTCGGGCAAAGTCCGAACA 4203
          :::::::::::::::         |||:::|||::::::
2034 .....ValAlaIleGlyPheGlnAlaLysAlaAspGlyGluAlaAla... 2047

4204 CGCCGTCAATACCGCCGTATTGGCTCAGGATTTCGGCAAAACCCGCAGTGC 4253
               :::|||:::      |||      :::   ::::::|||
2048 .........ValAlaIleGlyArgGlnThrGlnAlaGlyAsnGlnSerIl 2061

4254 GGAATGGGGCGTAAACGCCGAAATCAAAGGTTTCACGCTGTCCCTCCACG 4303
           |||   |||||||:::      |||      :::
2061 eAlaIleGlyAspAsnAlaGlnAlaThrGly.........AspGlnSerI 2075

4304 CTGCCGCCGCCAAAGGCCCGCAACTGGAAGCGCAACACAGCGCGGGCATC 4353
       |||   :::   |||      :::   ::::::||||||:::::::|||
2075 leAlaIleGlyThrGlyAsnValValThrGlyLysHisSerGlyAlaIle 2091
```

```
seq_name: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-413-814-78

seq_documentation_block:
; Sequence 78, Application US/09413814
; Patent No. 6225064
; GENERAL INFORMATION:
;  APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
;  APPLICANT: Bristol-Myers Squibb, Co.
;  APPLICANT: Beyer, Stefan
;  APPLICANT: Bloecker, Helmut
;  APPLICANT: Brandt, Petra
;  APPLICANT: Cino, Paul M
;  APPLICANT: Dougherty, Brian A
;  APPLICANT: Goldberg, Steven L
;  APPLICANT: Hofle, Gerhard
;  APPLICANT: Mueller, Joachim
;  APPLICANT: Reichenbach, Hans
;  TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polyketide or
;  TITLE OF INVENTION: heteropolyketide compounds
;  FILE REFERENCE: PCT/US 99/23535
;  CURRENT APPLICATION NUMBER: US/09/413,814
;  CURRENT FILING DATE: 1999-10-07
;  EARLIER APPLICATION NUMBER: DE 198 46 493.2
;  EARLIER FILING DATE: 1998-10-09
;  NUMBER OF SEQ ID NOS: 107
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 78
;   LENGTH: 882
;   TYPE: PRT
;   ORGANISM: Sorangium cellulosum
US-09-413-814-78


alignment_scores:
            Quality:  231.50            Length:    814
```

```
            Ratio:  0.693                Gaps:     35
 Percent Similarity:  41.032   Percent Identity: 25.061

alignment_block:
 US-09-303-518D-649 x US-09-413-814-78   ..

 Align seg 1/1  to: US-09-413-814-78  from: 1  to: 882

    2419 CAAGCAACATTTAATCAAGCCACATTAAACGGCAACACATCGGCTTCGGG 2468
         :::|||:::|||:::::::|||   :::|||:::   :::::::|||:::
       9 GluAlaAlaPheArgSerAla...PheAsnAlaCysAlaAlaAlaAlaAr 24

    2469 CAATGCTTCATTTAATCTAAGCGACCACGCCGTACAAAACGGC....... 2511
              :::::::::|||:::::   :::   :::::::|||
      24 gAlaArgGlyLeuAspLeuGlyGluValValPheArgGluGlyGluGlyA 41

    2512 ..AGTCTGACGCTTTCCGGCAACGCTAAGGCAAAC.............. 2544
           :::   ||||||   |||:::|||   |||:::
      41 spGlyArgThrLeuLeuGlyThrAlaLeuAlaGlnProAlaLeuPheAla 57

    2544 ............................................... 2544

      58 ValGluLeuAlaLeuAlaArgLeuTrpMetSerTrpGlyIleGluProAl 74

    2545 .......GTAAGCCATTCC.....................GCACTCAACG 2566
                 ::::::|||||                    ::::::   |
      74 aAlaMetIleGlyHisSerLeuGlyGluLeuValAlaAlaCysValAlaG 91

    2567 GTAATGTCTCCCTAGCCGATAAGGCAGTATTCCATTTTGAAAGCAGCCGC 2616
         ||      :::|||   |||          :::   ::::::   :::|||
      91 lyValPheThrLeuGluAspAlaMetSerLeuValIleAspArgGlyArg 107

    2617 TTTACCGGACAAATCAGCGGCGGCAAGGATACGGCATTACACTTA..... 2661
         |||                :::|||:::      |||:::   |||
     108 PheMetGlnAlaAlaProAlaGlySerMetLeuAlaValGlyLeuProAl 124

    2662 .AAAGACAGCGAATGGACGCTGCCGTCAGGCACGGAATTAGGCAATTTAA 2710
            |||   |||      |||   :::|||   |||::::::   :::|
     124 aAlaAspValGluGlyLeuLeuGluAlaGlyLeuGluIleAlaAlaValA 141

    2711 ACCTTGACAACGCCACCATTACACTCAAT......TCCGCCTATCGCCAC 2754
         ||      :::   |||::::::   :::      ||||||   |||
     141 snSerProLysLeuThrValValAlaGlyProAlaSerAlaIleArgAsp 157

    2755 GATGCGGCAGGGGCGCAAACCGGCAGTGCGACAGATGCGCCGCGCCGCCG 2804
            ||||||      ::::::   ::::::          |||   :::
     158 LeuAlaAlaArgLeuGluAlaArgGluValPheAlaArgProLeuGlnTh 174

    2805 TTCG................................................ 2808
          |||
     174 rSerHisAlaPheHisCysAlaLeuIleAspGlyAlaValAlaProPheL 191

    2809 ..........CGCCGTTCGCGCCG.............TTCCCTATTATC 2834
                  ||||||:::|||               :::|||::::::
     191 euGluSerValArgArgAlaArgLeuSerProProGluIleProValVal 207

    2835 ..................................................CG 2836

     208 SerAsnValThrGlyAlaLeuLeuThrAspAlaGluAlaThrAspProAl 224

    2837 TTACACCGCCAACTTCGGTAGAATCCCGTTTCAACACGCTGACGGTAAAC 2886
          |||   |||:::      |||   |||   :::   :::||||||   :
     224 aTyrTrpAlaArgHisLeuArgGlnProValArgPheSerAspGlyValG 241

    2887 GGCAAATTGAACGGTCAGGGAAC...ATTCCGCTTTATGTCGGAACT... 2930
         ::   :::      ||||||:::   :::   |||   ||||||
     241 luAlaLeuPheAlaSerGlyHisAlaLeuPheLeuGluValGlyProGly 257

    2931 ......CTTCGGCTACCGCAGCGACAAATTGAAGCTGGCGGAAAGTTCCG 2974
               |||           ||||||      ||||||  :::::::::
     258 ArgGlyLeuThrThrLeuValArgGlnThrLeuAlaGly.ArgGlyGlyA 274

    2975 AAGGCACTTACACCTTGGCGGTCAACAATACCGGCAACGAACCTGCAAGC 3024
```

```
         :::      :::|||:::   ::::::::::::::|||||||||||
     274 laAlaIleAlaSerLeuGlySerThrHisAlaAlaSerGluProAlaSer 290

    3025 CTCGAACAATTGACGGT.............AGTGGAAGGAAAAGACAACA 3061
         |||   :::                     :::|||:::      :::
     291 LeuAlaGluAlaLeuGlyGlnLeuTrpGluAlaGlyHisAlaValAspTr 307

    3062 AACCGCTGTCCGAAAACCTTAATTTCACCCTGCAAAACGAACACGTCGAT 3111
          ||||||      :::|||      :::|||      |||:::      |
     307 pThrAla......ArgProArgGlyArgProProAlaArgSerSerAlaC 322

    3112 GCCGGCGCGTG.........GCGTTACCAACTCATCCGCAAAGACGGCGA 3152
         ||||||||:::         :::      |||      :::|||
     322 ysArgArgThrArgSerSerGlyArgGlyThrGlySerArgArgAlaAla 338

    3153 GTTCCGCCTGCATAATCCGGTCAAAGAACAAGAGCTTTCCGACAAACTCG 3202
         :::|||||||      |||      |||   |||      |||   :::||
     339 AlaProProProProSer......ArgArgArgProThrArgProSerAr 353

    3203 GCA.AGGCAGAAGCCAAAAAACAGGCGGAAAAAGACAACGCGCAAAGCCT 3251
         |   |||:::   |||::::::||||||::::::   :::|||
     353 gProArgArgThrProSerArgArgArgArgSerProAlaArgThrProG 370

    3252 TGACGCGCTGATTGCGGCCGGGCGCGATGCCGTCGAAAAGACAGAAAGCG 3301
               |||      |||            ||||||   |||:::   |||
     370 lyProArgSerProArgArgArgProSerProSerAlaArgSerProAla 386

    3302 TTGCCGAACCGGCCCGGCAGGCAGGCGGGGAAAATGTCGGCATTATGCAG 3351
         :::   :::||||||
     387 IleTrpGluArgPro.....................................Pr 392

    3352 GCGGAGGAAGAGAAAAAACGGGTGCAGGCGGATAAAGACACCGCCTTGGC 3401
          ||||||   |||   :::      ||||||:::      |||      |
     392 oArgArgProArgAspArgArgProArgArgLeuLeuArgProAla...A 408

    3402 GAAACAGCGCGAAGCGGAAACCCGGCCGGCTACCACCGCCTTCCCCCGCG 3451
         ||:::   |||:::|||   ||||||   :::|||   |||      |||
     408 rgArgLeuAlaArgArgAspProGlyAspValProAspProArgAlaAla 424

    3452 CCCGCCGCGCCCGCCGGGATTTGCCGCAACTGCAACCCCAACCGCAGCCC 3501
               |||   |||      ||||||   |||   |||   |||
     425 ArgArg.ArgAlaPro......AlaAlaGlyAlaLeuProGlyAlaAspG 439

    3502 CAACCGCAGCGCGACCTGATCAGCCGTTATGCCAATAGCGGTTTGAGTGA 3551
         :::|||
     439 lyArgGlyAla..................................... 442

    3552 ATTTTCCGCCACGCTCAACAGCGTTTTCGCCGTACAGGACGAATTAGACC 3601
                                  |||      ||||||
     443 ..........................ArgGlyProGlyArg........ 447

    3602 GCGTATTTGCCGAAGACCGCCGCAACGCCGTTTGGACAAGCGGCATCCGG 3651
               ||||||   |||:::|||            |||   |||
     448 ..........ArgArgAlaProArgArgGlyAlaAlaAlaArgValProP 461

    3652 GACACCAAACACTACCGTTCGCAAGATTTCCGCGCCTACCGCCAACAAAC 3701
                                       |||      |||   :::
     461 ro..........................AspArgAlaAspProProArg 468

    3702 CGACCTGCGCCAAATCGGTATGCAGAAAAACCT.........CGGCAGCG 3742
         ||||||||||||                |||          |||:::
     469 ArgProAlaProAlaValValArgAlaAlaProAlaValValArgArgSe 485

    3743 GGCGCGT..................CGGCATCCTGTTTTCGCACAACCGG 3774
          ||||||                 :::|||||||||   :::      |
     485 rAlaArgAlaGlyGlnProGlyValGlnHisProValArgGlyAlaArgA 502

    3775 ACCGAAAACACCTTCGACGACGG........................... 3797
         ||           |||||||||
     502 spGlyProAlaArgArgArgArgAlaProProGlnProValArgAspArg 518

    3798 ............................CATCGGCAACTCGGCACGGC 3817
                                     :::|||||||         ||
```

```
     519 AlaAlaProArgGlyAlaAlaAspAspValGlnArgGlnGlyArgArgAl 535

    3818 TTGCCCACGGCGCCGTTTTCGGGCA.................ATACGGC 3849
         |   |||   ||||||   |||:::                 :::|||
     535 aValProGlyArgArgAlaArgGlyProGlyAlaValProAspValArgP 552

    3850 ATCGACAGGTTCTACATCGGCATCAGCGCGGGCGCGGGTTTTAGCAGCGG 3899
           |||          |||      |||   ||||||
     552 roArgAlaProArgGlyArgGlyProArg...ArgGlyGlyLeuGlyAla 567

    3900 CAGCCTTTCAGACGGCATCGGAGGCAAAATCCGCCGCCGCGTGCTGCATT 3949
         :::|||      :::   |||      :::|||   ||||||||||||
     568 ArgProArgGlyGluProArgAlaValArgProGluProArgAlaAlaPr 584

    3950 ACGGCATTCAGGC..................................... 3962
             ::::::|||
     584 oProArgAlaGlyHisProGluAlaProArgArgAlaArgHisArgAlaA 601

    3963 ..ACGATACCGCGCCGGTTTCGGCGGATTCGGCATCGAACCGCACATCGG 4010
              |||||||||   ||||||:::      |||      :::|||
     601 rgArgAlaProArgArgValArgArgLeuValGlyArgArgLeuArgArg 617

    4011 ......CGCAACGCGCTATTTCGT.....................CCAAA 4033
               |||:::||||||   |||                      |||
     618 ArgAlaArgArgAlaLeuArgArgLeuArgAlaGlyProAlaPheProAl 634

    4034 AAGCGGATTACCGCTACGAAAACGTCAATATCGCCACCCCCGGCCTTGCA 4083
         :::|||   |||         |||:::   |||:::      |||:::
     634 aAlaGlyAlaProGlyAlaValArgArgLeuArgArg...SerProAlaG 650

    4084 TTCAACCGCTACCGCGCGGGCATTAAGGCAGATTATTCATTCAAACCGGC 4133
                 :::   ||||||      ||||||
     650 lyValAlaValArgArgGlyProGlyGlyArg..................... 660

    4134 GCAACACATTTCCATCACGCCTTATTTGAGCCTGTCCTATACCGATGCCG 4183
                                      |||||||||
     661 ............................AlaProValLeuAspAspGluAl 668

    4184 CTTCGGGCAAAGTCCGAACACGCGTCAATACCGCCGTATTGGCTCAGGAT 4233
         :::|||::::::            :::      |||      :::|||
     668 aLeuGlyArgAlaAlaProGlyProArgAlaGlyArgProArgAlaGlyG 685

    4234 TTCGGCAAAACCCGCAGTGCGGAATGGGGCGTAAACGCCGAAATCAAAGG 4283
              ::::::         |||   ::::::   |||||||:::   |||
     685 lyAlaLysAspValAlaArgGlyAlaAlaGluProArgArgArgGlyArg 701

    4284 TTTCACGCTGTC...................................... 4295
            |||
     702 AlaHisProProAspGlnGlyValLeuArgAlaGlyGlyAspLeuProLe 718

    4296 .......CCTCCACGCTGCCGCCGCCAAAGGCCCGCA.........ACTG 4329
                 |||   |||      :::   |||||||||          :
     718 uHisGlyProAlaArgArgValGlnAlaArgProAlaProAlaHisArgA 735

    4330 GAAGCGCAACACAGCGCGGGCATCAAATTAGGCTACCG 4367
         ::   :::   :::||||||||:::   |||   |||
     735 laArgGlyProArgArgGlyHisArgArgArgGluPro 747

seq_name: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-377-155-33

seq_documentation_block:
; Sequence 33, Application US/09377155
; Patent No. 6197312
; GENERAL INFORMATION:
;  APPLICANT:  PEAK, Ian Richard Anselm
;  APPLICANT:  JENNINGS, Michael Paul
;  APPLICANT:  MOXON, E. Richard
;  TITLE OF INVENTION:  NOVEL SURFACE ANTIGEN
;  FILE REFERENCE:  065064/0128
;  CURRENT APPLICATION NUMBER: US/09/377,155
;  CURRENT FILING DATE:  1999-08-19
;  PRIOR APPLICATION NUMBER:  PCT/AU98/01031
;  PRIOR FILING DATE:  1998-12-14
;  PRIOR APPLICATION NUMBER:  GB 9726398.2
;  PRIOR FILING DATE:  1997-12-12
;  NUMBER OF SEQ ID NOS:  33
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 33
;   LENGTH: 2353
;   TYPE: PRT
;   ORGANISM: Haemophilus influenzae
US-09-377-155-33


alignment_scores:
            Quality:  229.50              Length:   1495
              Ratio:  0.339                 Gaps:     72
 Percent Similarity:  45.284   Percent Identity: 19.130

alignment_block:
 US-09-303-518D-649 x US-09-377-155-33   ..

 Align seg 1/1  to: US-09-377-155-33  from: 1  to: 2353

     172 GCCGAAAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTA 221
         |||::::::   |||   :::          :::||||||   ||||||
     436 AlaAspThrAspGlyAlaLeuGluGlyIleSerLysAspGlnGluVal.. 451

     222 CAACAAAAAAGGGGAGTTGGTC.............GGCAAATCAATG...A 256
             |||   ||||||   |||             ||||||::::::   :
     452 ....LysAlaGlyGluThrValThrPheLysAlaGlyLysAsnLeuLysV 467

     257 CAAAAGCCCCGATGATTGATTTTTCTGTGGTGTCGCGTAACGGCGTGGCG 306
         ::|||             :::|||:::          ::::::::::::::
     467 alLysGlnAspGlyAlaAsnPheThrTyrSerLeuGlnAspAlaLeuThr 483

     307 GCATTGGTGGGC...................................... 318
         :::|||::::::
     484 GlyLeuThrSerIleThrLeuGlyGlyThrThrAsnGlyGlyAsnAspAl 500

     319 ...................GATCAATATATTGTGAGCGTGGCACATAACG 349
                             |||         ::::::   |||   ||||
     500 aLysThrValIleAsnLysAspGlyLeuThrIleThrProAlaGlyAsnG 517

     350 GCGGCTATAACAACGTTGATTTTGGTGCGGAAGGAAGAAATCCCGATCAA 399
         ||                      |||:::   |||   |||
     517 ly.....................GlyThrThrGlyThrAsn......... 523

     400 CATCGTTTTACTTATAAAATTGTGAAACGGAATAATTATAAAGCAGGGAC 449
                 |||   ::::::::::|||   ::::::   |||||||||::
     524 .........ThrIleSerValThrLys...AspGlyIleLysAlaGlyAs 536

     450 TAAAGGCCATCCTTATGGCGGCGATTATCATATGCCGCGTTTGCATAAAT 499
         :|||:::
     536 nLysAla................................................ 538

     500 TTGTCACAGATGCAGAACCTGTTGAAATGACCAGTTATATGGATGGGCGG 549
                                       :::|||:::      :::|||
     539 .........................IleThrAsnValAlaSerGlyLeu 546

     550 AAATATATCGATCAAAATAATTACCCTGACCGTGTTCGTATTGGGGCAGG 599
         :::       |||:::   |||:::
     547 ArgAlaTyrAspAspAlaAsnPhe........................... 554

     600 CAGGCAATATTGGCGATCTGATGAAGATGAGCCCAATAAC......CGCG 643
                                 |||      ||||||         :
     555 .......................AspValLeuAsnAsnSerAlaThrA 563

     644 AAAGTTCATATCATATTGCAAGTGCGTATTCTTGGCTCGTTGGTGGCAAT 693
         ::   :::   |||:::   :::|||||||:::
     563 spLeuAsnArgHisValGluAspAlaTyrLys.................. 573

     694 ACCTTTGCACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGTGA 743
                                     |||   :::||||||:::::::
     574 .........................GlyLeuLeuAsnLeuAsnGluLy 581
```

```
 744 AAAAATTAAACATAGCCCATATGGTTTTTTACCAACAGGAGGCTCATTTG 793
581 sAsnAlaAsnLysGlnProLeuValThrAspSerThrAlaAlaThrValG 598

 794 GCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGG 843
598 ly..........................AspLeuArgLysLeuGlyTrp 605

 844 TTAATT.........AATGGGGTATTGCAAACGGGCAACCCCTATATAGG 884
606 ValValSerThrLysAsnGlyThrLysGluGluSerAsn........... 618

 885 AAAAAGCAATGGCTTCCAGCTGGTTCGTAAAGATTGGTTCTATGATGAAA 934
619 ................GlnValLysGlnAlaAsp.........GluValL 627

 935 TCTTTGCTGGAGATACCCATTCAGTATTCTACGAACCACGTCAAAATGGG 984
627 euPheThrGlyAlaGlyAlaAlaThrValThrSerLysSerGluAsnGly 643

 985 AAATACTCTTTTAACGACGATAATAATGGCACAGGAAAAATCAATGCCAA 1034
644 LysHisThr..................................... 646

1035 ACATGAACACAATTCTCTGCCTAATAGATTAAAAACACGAACCGTTCAAT 1084

 646 ..................................................... 646

1085 TGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTGTTTATCATGCT 1134
647 ..IleThrValSerValAlaGluThrLysAlaAspCysGlyLeuGluLys 662

1135 GCAGGTGGTGTCAACAGTTATCGACCCAGACTGAATAATGGAGAAAATAT 1184
663 AspGlyAspThrIleLysLeuLysValAspAsnGlnAsnThrAspAsnVa 679

1185 TTCCTTTATT.................................... 1194
679 lLeuThrValGlyAsnAsnGlyThrAlaValThrLysGlyGlyPheGluT 696

1195 ...................GACGAAGGAAAAGGCGAATTGATACTT... 1221
696 hrValLysThrGlyAlaThrAspAlaAspArgGlyLysValThrValLys 712

1222 ......ACCAGCAAC.............................. 1230
713 AspAlaThrAlaAsnAspAlaAspLysLysValAlaThrValLysAspVa 729

1231 ..........ATCAATCAAGGTGCTGGAGGATTATATTTCCAAGGAGATT 1270
729 lAlaThrAlaIleAsnSerAlaAlaThrPheValLysThrGluAsnLeuT 746

1271 TTACGGTCTCGCCTGAAAATAACGAAACTTGGCAAGGCGCGGGCGTTCAT 1320
746 hrThrSerIleAspGluAspAsnProThrAspAsnGlyLysAspAspAla 762

1321 ATCAGTGAAGACAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGA 1370
763 LeuLysAlaGlyAspThrLeuThrPheLysAlaGly......LysAsnLe 777

1371 CCGCCTGTCCAAAATCGGCAAAGGC..................... 1395
777 uLysValLysArgAspGlyLysAsnIleThrPheAspLeuAlaLysAsnL 794

1396 ..........................ACGCTGCACGTTCAAGCCAAA 1416
794 euGluValLysThrAlaLysValSerAspThrLeuThrIleGlyGlyAsn 810

1417 GGGGAAAACCAAGGCTCG.....................ATCAGCGT 1442
811 ThrProThrGlyGlyThrThrAlaThrProLysValAsnIleThrSerTh 827

1443 GGGCGACGGTACAGTCATTTTGGATCAGCAGGCAGACGATAAAGGCAAAA 1492

 827 rAlaAspGlyLeuAsnPheAlaLysGluThrAlaAspAlaSerGlySerL 844

1493 AACAAGCCTTT.............................AGTGAAATC 1512
 844 ysAsnValTyrLeuLysGlyIleAlaThrThrLeuThrGluProSerAla 860

1513 GGCTTGGTCAGCGGCAGGGGTACGGTGCAACTGAATGCCGATAATCAGTT 1562
 861 GlyAlaLysSerSerHisValAspLeuAsnValAspAlaThrLysLysSe 877

1563 CAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGA...CGTTTGGATT 1609
 877 rAsnAlaAlaSerIleGluAspValLeuArgAlaGlyTrpAsnIleGlnG 894

1610 TAAACGGGCATTCGCTTTCGTTCCACCGTATTCAAAATACCGATGAAGGG 1659
 894 lyAsnGlyAsnAsnValAspTyrValAlaThrTyrAspThr......... 907

1660 GCGATGATTGTCAACCACAATCAAGACAAAGAATCCACCGTTACCATTAC 1709
 908 .........ValAsnPheThrAspAspSerThrGlyThrThrThrValTh 921

1710 AGGCAATAAAGATATTGCTACAACCGGCAATAACAACAGCTTGGATAGCA 1759
 921 rValThrGlnLysAlaAspGlyLysGlyAlaAspValLysIleGlyAlaL 938

1760 AAAAAGAAATTGCC......TACAACGGT...TGGTTTGGCGAGAAAGAT 1800
 938 ysThrSerValIleLysAspHisAsnGlyLysLeuPheThrGlyLysAsp 954

1801 ACGACCAAAACGAACGGGCGGCTCAACCTTGTTTACCAGCCCGCCGCAGA 1850
 955 LeuLysAspAlaAsnAsnGlyAlaThrValSerGluAspAspGlyLysAs 971

1851 AGACCGCACCCTGCTGCTTTCCGGCGGAACA..............AATT 1885
 971 pThrGlyThrGlyLeuValThrAlaLysThrValIleAspAlaValAsnL 988

1886 TAAACGGCAACATCACGCAAACAAACGGCAAACTGTTTTTCAGCGGCAGA 1935
 988 ysSerGlyTrpArgValThrGlyGluGlyAlaThrAlaGluThrGlyAla 1004

1936 CCAACACCGCACGCCTACAATCATTTAAACGACCATTGGTCGCAAAAAGA 1985
1005 ThrAlaValAsnAlaGlyAsn.......................AlaGl 1013

1986 GGGCATTCCTCGCGGGGAAATCGTGTGGGACAACGACTGGATCAACCGCA 2035
1013 uThrValThrSerGlyThrSerVal..................... 1021

2036 CATTTAAAGCGGAAAACTTCCAAATTAAAGGCGGACAGGCGGTGGTTTCC 2085
1022 ..............AsnPhe......LysAsnGlyAsnAlaThrThrAla 1031

2086 CGCAATGTTGCCAAAGTGAAAGGCGATTGGCATTTGAGCAATCACGCCCA 2135
1032 ...ThrValSerLysAspAsnGlyAsnIleAsnValLysTyrAspValAs 1047

2136 AGCAGTTTTTGGTGTCGCACCGCATCAAAGCCACACAATCTGTACACGTT 2185
1047 nValGlyAspGlyLeuLysIleGlyAspAspLysLysIleVal......A 1062

2186 CGGACTGGACGGGTCTGACAAATTGTGTCGAAAAAACCATTACCGACGAT 2235
1062 laAspThrThrThrLeu.................ThrValThrGlyGly 1072

2236 AAAGTGATTGCTTCATTGACTAAGACCGACATCAGCGGCAATGTCGATCT 2285
1073 LysValSerValProAlaGlyAlaAsnSerValAsnAsnAsnLysLysLe 1089

2286 TGCCGATCACGCTCATTTAAATCTCACAGGGCTTGCCACACTCAACGGCA 2335
```



```
    3775 ..........ACCGAAAACACCTTCGACGACGGCATCGGCAACTCGGCA 3813
                   |||:::  ||||||  ||||||  |||::::::|||
    1648 alAlaSerGlyThrAsnValThrPheAlaAspGlyAsnGlyThrThrAla 1664

    3814 CGGCTTGCC.........CACGGCGCCGTTTTCGGGCAATACGGCATCGA 3854
         :::::::::          |||::::::      :::|||::::::
    1665 GluValThrLysAlaAsnAspGlySerIleThrValLysTyrAsnVal.. 1680

    3855 CAGGTTCTACATCGGCATCAGCGCGGGCGCGGGTTTTAGCAGCGGCAGCC 3904
                                                      ::::
    1681 .....................................................LysV 1682

    3905 TTTCAGACGGCATC.........GGAGGCAAAATC 3930
         :::::|||||||:::         |||   ||||||
    1682 alAlaAspGlyLeuLysLeuAspGlyAspLysIle 1693

seq_name: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-913-942-4

seq_documentation_block:
; Sequence 4, Application US/08913942
; Patent No. 6200578
;  GENERAL INFORMATION:
;    APPLICANT:  St. Geme, Joseph
;    APPLICANT:  Barenkamp, Stephen J.
;    TITLE OF INVENTION:  HAEMOPHILUS ADHESION PROTEINS
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Flehr Hohbach Test Albritton & Herbert LLP
;      STREET:  Four Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/913,942
;      FILING DATE:  29-DEC-1997
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/409,995
;      FILING DATE:  24-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US96/4031
;      FILING DATE:  22-MAR-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Vance, Dolly A.
;      REGISTRATION NUMBER:  39,054
;      REFERENCE/DOCKET NUMBER:  A-61053-1/RFT/RMS/DAV
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2353 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-913-942-4


alignment_scores:
           Quality:  229.50            Length:   1495
             Ratio:  0.339               Gaps:     72
 Percent Similarity:  45.284   Percent Identity: 19.130

alignment_block:
 US-09-303-518D-649 x US-08-913-942-4    ..
```

```
Align seg 1/1  to: US-08-913-942-4  from: 1  to: 2353

    172 GCCGAAAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTA 221
        |||::::::  |||  :::         :::||||||  ||||||
    436 AlaAspThrAspGlyAlaLeuGluGlyIleSerLysAspGlnGluVal.. 451

    222 CAACAAAAAAGGGGAGTTGGTC.............GGCAAATCAATG...A 256
            |||  ||||||  |||             ||||||::::::   :
    452 ....LysAlaGlyGluThrValThrPheLysAlaGlyLysAsnLeuLysV 467

    257 CAAAAGCCCCGATGATTGATTTTTCTGTGGTGTCGCGTAACGGCGTGGCG 306
        ::|||            :::|||:::         ::::::::::::::
    467 alLysGlnAspGlyAlaAsnPheThrTyrSerLeuGlnAspAlaLeuThr 483

    307 GCATTGGTGGGC..................................... 318
        :::|||::::::
    484 GlyLeuThrSerIleThrLeuGlyGlyThrThrAsnGlyGlyAsnAspAl 500

    319 ...................GATCAATATATTGTGAGCGTGGCACATAACG 349
                           |||         ::::::  |||  ||||
    500 aLysThrValIleAsnLysAspGlyLeuThrIleThrProAlaGlyAsnG 517

    350 GCGGCTATAACAACGTTGATTTTGGTGCGGAAGGAAGAAATCCCGATCAA 399
        ||                    |||:::  |||  |||
    517 ly.....................GlyThrThrGlyThrAsn......... 523

    400 CATCGTTTTACTTATAAAATTGTGAAACGGAATAATTATAAAGCAGGGAC 449
              |||  :::::::::|||  ::::::  |||||||||::
    524 .........ThrIleSerValThrLys...AspGlyIleLysAlaGlyAs 536

    450 TAAAGGCCATCCTTATGGCGGCGATTATCATATGCCGCGTTTGCATAAAT 499
        :|||:::'
    536 nLysAla..................................... 538

    500 TTGTCACAGATGCAGAACCTGTTGAAATGACCAGTTATATGGATGGGCGG 549
                                  :::|||:::      :::|||
    539 .........................IleThrAsnValAlaSerGlyLeu 546

    550 AAATATATCGATCAAAATAATTACCCTGACCGTGTTCGTATTGGGGCAGG 599
        :::      |||:::  |||:::
    547 ArgAlaTyrAspAspAlaAsnPhe..................... 554

    600 CAGGCAATATTGGCGATCTGATGAAGATGAGCCCAATAAC......CGCG 643
                                  |||       ||||||       :
    555 .........................AspValLeuAsnAsnSerAlaThrA 563

    644 AAAGTTCATATCATATTGCAAGTGCGTATTCTTGGCTCGTTGGTGGCAAT 693
        ::   :::  |||:::  :::|||||||:::
    563 spLeuAsnArgHisValGluAspAlaTyrLys.................. 573

    694 ACCTTTGCACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGTGA 743
                                     |||  :::||||||:::::::
    574 .........................GlyLeuLeuAsnLeuAsnGluLy 581

    744 AAAAATTAAACATAGCCCATATGGTTTTTTACCAACAGGAGGCTCATTTG 793
        ::::  :::  :::|||                |||::::::::::   |
    581 sAsnAlaAsnLysGlnProLeuValThrAspSerThrAlaAlaThrValG 598

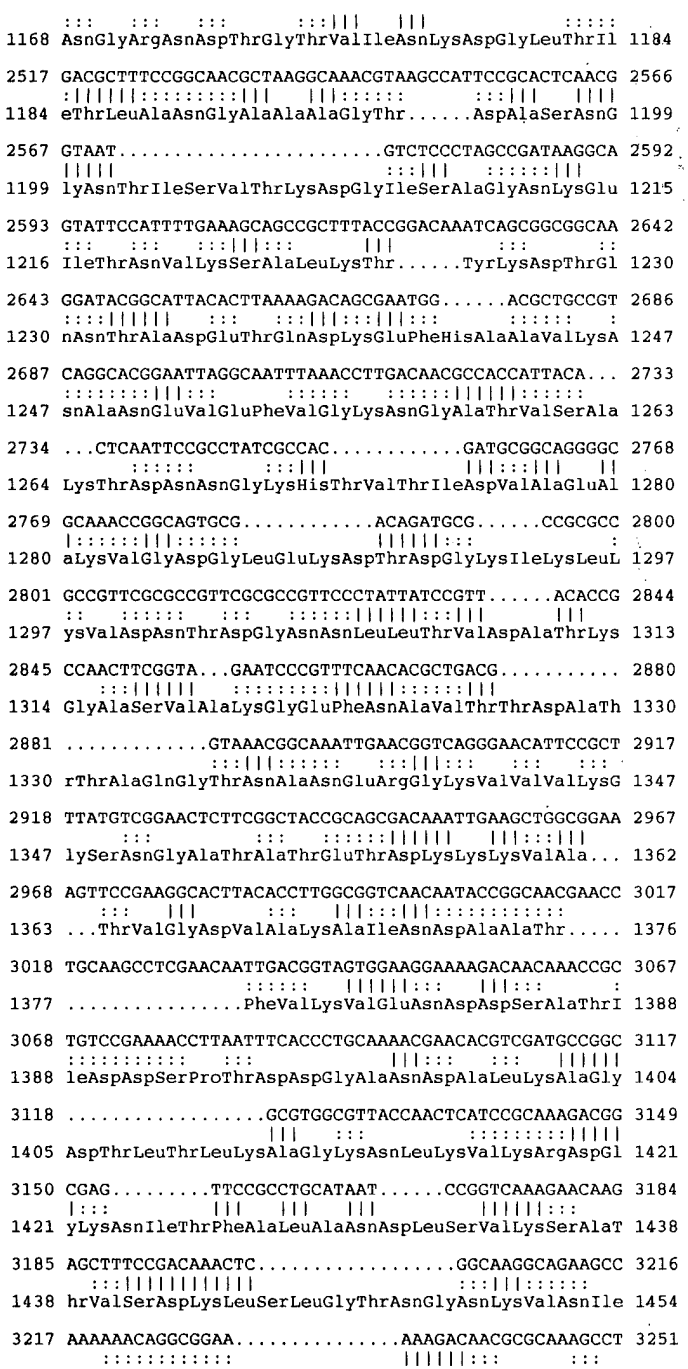
    794 GCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGG 843
        ||                             |||  :::|||      |||
    598 ly.........................AspLeuArgLysLeuGlyTrp 605

    844 TTAATT.........AATGGGGTATTGCAAACGGGCAACCCCTATATAGG 884
        ::::::         ||||||:::  :::  :::|||
    606 ValValSerThrLysAsnGlyThrLysGluGluSerAsn........... 618

    885 AAAAAGCAATGGCTTCCAGCTGGTTCGTAAAGATTGGTTCTATGATGAAA 934
                      |||:::  :::  |||         :::  :
    619 ..............GlnValLysGlnAlaAsp.........GluValL 627

    935 TCTTTGCTGGAGATACCCATTCAGTATTCTACGAACCACGTCAAAATGGG 984
        ::|||:::|||         ::::::      :::      :::||||||
    627 euPheThrGlyAlaGlyAlaAlaThrValThrSerLysSerGluAsnGly 643
```







seq_name: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-669-974-33

seq_documentation_block:

```
; Sequence 33, Application US/09669974
; Patent No. 6333173
; GENERAL INFORMATION:
;  APPLICANT: PEAK, Ian Richard Anselm
;  APPLICANT:  JENNINGS, Michael Paul
;  APPLICANT:  MOXON, E. Richard
;  TITLE OF INVENTION: NOVEL SURFACE ANTIGEN
;  FILE REFERENCE: 065064/0128
;  CURRENT APPLICATION NUMBER: US/09/669,974
;  CURRENT FILING DATE:  2000-09-26
;  PRIOR APPLICATION NUMBER: US 09/377,155
;  PRIOR FILING DATE: 1999-08-19
;  PRIOR APPLICATION NUMBER: PCT/AU98/01031
;  PRIOR FILING DATE: 1998-12-14
;  PRIOR APPLICATION NUMBER: GB 9726398.2
;  PRIOR FILING DATE: 1997-12-12
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 33
;   LENGTH: 2353
;   TYPE: PRT
;   ORGANISM: Haemophilus influenzae
US-09-669-974-33


alignment_scores:
            Quality:  229.50              Length:   1495
              Ratio:  0.339                Gaps:     72
 Percent Similarity:  45.284   Percent Identity: 19.130

alignment_block:
 US-09-303-518D-649 x US-09-669-974-33   ..

 Align seg 1/1  to: US-09-669-974-33  from: 1  to: 2353

     172 GCCGAAAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTA 221
         |||::::::   |||   :::          :::||||||   ||||||
     436 AlaAspThrAspGlyAlaLeuGluGlyIleSerLysAspGlnGluVal.. 451

     222 CAACAAAAAAGGGGAGTTGGTC............GGCAAATCAATG...A 256
             |||   ||||||   |||             ||||||::::::   :
     452 ....LysAlaGlyGluThrValThrPheLysAlaGlyLysAsnLeuLysV 467

     257 CAAAAGCCCCGATGATTGATTTTTCTGTGGTGTCGCGTAACGGCGTGGCG 306
         ::|||             :::|||:::          ::::::::::::::::
     467 alLysGlnAspGlyAlaAsnPheThrTyrSerLeuGlnAspAlaLeuThr 483

     307 GCATTGGTGGGC..................................... 318
         :::|||::::::
     484 GlyLeuThrSerIleThrLeuGlyGlyThrThrAsnGlyGlyAsnAspAl 500

     319 ...................GATCAATATATTGTGAGCGTGGCACATAACG 349
                           |||          ::::::   |||   ||||
     500 aLysThrValIleAsnLysAspGlyLeuThrIleThrProAlaGlyAsnG 517

     350 GCGGCTATAACAACGTTGATTTTGGTGCGGAAGGAAGAAATCCCGATCAA 399
         ||                        |||:::   |||   |||
     517 ly....................GlyThrThrGlyThrAsn......... 523

     400 CATCGTTTTACTTATAAAATTGTGAAACGGAATAATTATAAAGCAGGGAC 449
                 |||   ::::::::::|||   ::::::   |||||||||::
     524 .........ThrIleSerValThrLys...AspGlyIleLysAlaGlyAs 536

     450 TAAAGGCCATCCTTATGGCGGCGATTATCATATGCCGCGTTTGCATAAAT 499
         :|||:::
     536 nLysAla........................................... 538

     500 TTGTCACAGATGCAGAACCTGTTGAAATGACCAGTTATATGGATGGGCGG 549
                                    :::|||:::      :::|||
     539 ........................IleThrAsnValAlaSerGlyLeu 546

     550 AAATATATCGATCAAAATAATTACCCTGACCGTGTTCGTATTGGGGCAGG 599
         :::      |||:::   |||:::

     547 ArgAlaTyrAspAspAlaAsnPhe........................... 554

     600 CAGGCAATATTGGCGATCTGATGAAGATGAGCCCAATAAC......CGCG 643
                                    |||      ||||||      :
     555 ........................AspValLeuAsnAsnSerAlaThrA 563

     644 AAAGTTCATATCATATTGCAAGTGCGTATTCTTGGCTCGTTGGTGGCAAT 693
         ::   :::   |||:::   :::||||||:::
     563 spLeuAsnArgHisValGluAspAlaTyrLys................. 573

     694 ACCTTTGCACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGTGA 743
                                   |||   :::||||||::::::::
     574 ..........................GlyLeuLeuAsnLeuAsnGluLy 581

     744 AAAAATTAAACATAGCCCATATGGTTTTTTACCAACAGGAGGCTCATTTG 793
         ::::   :::   :::|||                |||::::::::::   |
     581 sAsnAlaAsnLysGlnProLeuValThrAspSerThrAlaAlaThrValG 598

     794 GCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGG 843
         ||                              |||   :::|||      |||
     598 ly..........................AspLeuArgLysLeuGlyTrp 605

     844 TTAATT.........AATGGGGTATTGCAAACGGGCAACCCCTATATAGG 884
         ::::::         ||||||:::   :::   :::|||
     606 ValValSerThrLysAsnGlyThrLysGluGluSerAsn........... 618

     885 AAAAAGCAATGGCTTCCAGCTGGTTCGTAAAGATTGGTTCTATGATGAAA 934
                         |||:::   :::   |||          :::   :
     619 ...............GlnValLysGlnAlaAsp.........GluValL 627

     935 TCTTTGCTGGAGATACCCATTCAGTATTCTACGAACCACGTCAAAATGGG 984
         ::|||:::|||         ::::::      :::      :::|||||||
     627 euPheThrGlyAlaGlyAlaAlaThrValThrSerLysSerGluAsnGly 643

     985 AAATACTCTTTTAACGACGATAATAATGGCACAGGAAAAATCAATGCCAA 1034
         |||::::::
     644 LysHisThr....................................... 646

    1035 ACATGAACACAATTCTCTGCCTAATAGATTAAAAACACGAACCGTTCAAT 1084

     646 ..................................................... 646

    1085 TGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTGTTTATCATGCT 1134
           ::::::||||||::::::||||||      :::      :::
     647 ..IleThrValSerValAlaGluThrLysAlaAspCysGlyLeuGluLys 662

    1135 GCAGGTGGTGTCAACAGTTATCGACCCAGACTGAATAATGGAGAAAATAT 1184
           |||   :::   :::   :::          :::|||   :::|||::
     663 AspGlyAspThrIleLysLeuLysValAspAsnGlnAsnThrAspAsnVa 679

    1185 TTCCTTTATT....................................... 1194
         :      :::
     679 lLeuThrValGlyAsnAsnGlyThrAlaValThrLysGlyGlyPheGluT 696

    1195 ....................GACGAAGGAAAAGGCGAATTGATACTT... 1221
                             |||      :::|||::::::   :::
     696 hrValLysThrGlyAlaThrAspAlaAspArgGlyLysValThrValLys 712

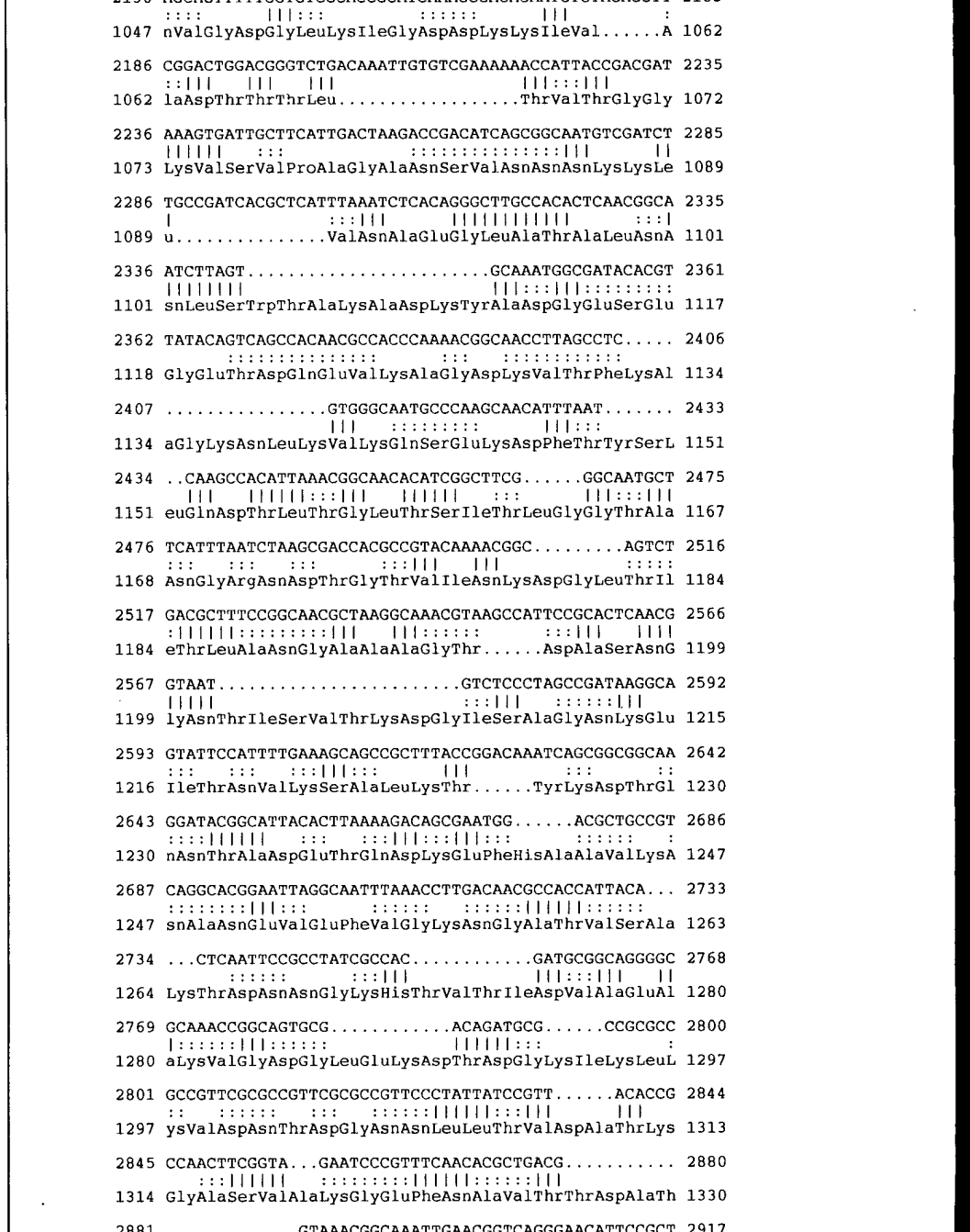
    1222 ......ACCAGCAAC................................... 1230
               |||:::|||
     713 AspAlaThrAlaAsnAspAlaAspLysLysValAlaThrValLysAspVa 729

    1231 ..........ATCAATCAAGGTGCTGGAGGATTATATTTCCAAGGAGATT 1270
                   ||||||::::::|||      :::      ::::::
     729 lAlaThrAlaIleAsnSerAlaAlaThrPheValLysThrGluAsnLeuT 746

    1271 TTACGGTCTCGCCTGAAAATAACGAAACTTGGCAAGGCGCGGGCGTTCAT 1320
           |||         |||:::|||   |||   :::|||
     746 hrThrSerIleAspGluAspAsnProThrAspAsnGlyLysAspAspAla 762

    1321 ATCAGTGAAGACAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGA 1370
         ::::::      :::|||:::|||:::|||::::::         |||
     763 LeuLysAlaGlyAspThrLeuThrPheLysAlaGly......LysAsnLe 777
```


1371 CCGCCTGTCCAAAATCGGCAAAGGC..... 1395
777 uLysValLysArgAspGlyLysAsnIleThrPheAspLeuAlaLysAsnL 794
1396 ACGCTGCACGTTCAAGCCAAA 1416
794 euGluValLysThrAlaLysValSerAspThrLeuThrIleGlyGlyAsn 810
1417 GGGGAAAACCAAGGCTCG.....ATCAGCGT 1442
811 ThrProThrGlyGlyThrThrAlaThrProLysValAsnIleThrSerTh 827
1443 GGGCGACGGTACAGTCATTTTGGATCAGCAGGCAGACGATAAAGGCAAAA 1492
827 rAlaAspGlyLeuAsnPheAlaLysGluThrAlaAspAlaSerGlySerL 844
1493 AACAAGCCTTT.....AGTGAAATC 1512
844 ysAsnValTyrLeuLysGlyIleAlaThrThrLeuThrGluProSerAla 860
1513 GGCTTGGTCAGCGGCAGGGGTACGGTGCAACTGAATGCCGATAATCAGTT 1562
861 GlyAlaLysSerSerHisValAspLeuAsnValAspAlaThrLysLysSe 877
1563 CAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGA...CGTTTGGATT 1609
877 rAsnAlaAlaSerIleGluAspValLeuArgAlaGlyTrpAsnIleGlnG 894
1610 TAAACGGGCATTCGCTTTCGTTCCACCGTATTCAAAATACCGATGAAGGG 1659
894 lyAsnGlyAsnAsnValAspTyrValAlaThrTyrAspThr..... 907
1660 GCGATGATTGTCAACCACAATCAAGACAAAGAATCCACCGTTACCATTAC 1709
908 ValAsnPheThrAspAspSerThrGlyThrThrThrValTh 921
1710 AGGCAATAAAGATATTGCTACAACCGGCAATAACAACAGCTTGGATAGCA 1759
921 rValThrGlnLysAlaAspGlyLysGlyAlaAspValLysIleGlyAlaL 938
1760 AAAAAGAAATTGCC.....TACAACGGT...TGGTTTGGCGAGAAAGAT 1800
938 ysThrSerValIleLysAspHisAsnGlyLysLeuPheThrGlyLysAsp 954
1801 ACGACCAAAACGAACGGGCGGCTCAACCTTGTTTACCAGCCCGCCGCAGA 1850
955 LeuLysAspAlaAsnAsnGlyAlaThrValSerGluAspAspGlyLysAs 971
1851 AGACCGCACCCTGCTGCTTTCCGGCGGAACA.....AATT 1885
971 pThrGlyThrGlyLeuValThrAlaLysThrValIleAspAlaValAsnL 988
1886 TAAACGGCAACATCACGCAAACAAACGGCAAACTGTTTTTCAGCGGCAGA 1935
988 ysSerGlyTrpArgValThrGlyGluGlyAlaThrAlaGluThrGlyAla 1004
1936 CCAACACCGCACGCCTACAATCATTTAAACGACCATTGGTCGCAAAAAGA 1985
1005 ThrAlaValAsnAlaGlyAsn.....AlaGl 1013
1986 GGGCATTCCTCGCGGGGAAATCGTGTGGGACAACGACTGGATCAACCGCA 2035
1013 uThrValThrSerGlyThrSerVal..... 1021
2036 CATTTAAAGCGGAAAACTTCCAAATTAAAGGCGGACAGGCGGTGGTTTCC 2085
1022 AsnPhe.....LysAsnGlyAsnAlaThrThrAla 1031
2086 CGCAATGTTGCCAAAGTGAAAGGCGATTGGCATTTGAGCAATCACGCCCA 2135
1032 ...ThrValSerLysAspAsnGlyAsnIleAsnValLysTyrAspValAs 1047


2136 AGCAGTTTTTGGTGTCGCACCGCATCAAAGCCACACAATCTGTACACGTT 2185
1047 nValGlyAspGlyLeuLysIleGlyAspAspLysLysIleVal.....A 1062
2186 CGGACTGGACGGGTCTGACAAATTGTGTCGAAAAAACCATTACCGACGAT 2235
1062 laAspThrThrThrLeu.....ThrValThrGlyGly 1072
2236 AAAGTGATTGCTTCATTGACTAAGACCGACATCAGCGGCAATGTCGATCT 2285
1073 LysValSerValProAlaGlyAlaAsnSerValAsnAsnAsnLysLysLe 1089
2286 TGCCGATCACGCTCATTTAAATCTCACAGGGCTTGCCACACTCAACGGCA 2335
1089 u.....ValAsnAlaGluGlyLeuAlaThrAlaLeuAsnA 1101
2336 ATCTTAGT.....GCAAATGGCGATACACGT 2361
1101 snLeuSerTrpThrAlaLysAlaAspLysTyrAlaAspGlyGluSerGlu 1117
2362 TATACAGTCAGCCACAACGCCACCCAAAACGGCAACCTTAGCCTC..... 2406
1118 GlyGluThrAspGlnGluValLysAlaGlyAspLysValThrPheLysAl 1134
2407 GTGGGCAATGCCCAAGCAACATTTAAT..... 2433
1134 aGlyLysAsnLeuLysValLysGlnSerGluLysAspPheThrTyrSerL 1151
2434 ..CAAGCCACATTAAACGGCAACACATCGGCTTCG.....GGCAATGCT 2475
1151 euGlnAspThrLeuThrGlyLeuThrSerIleThrLeuGlyGlyThrAla 1167
2476 TCATTTAATCTAAGCGACCACGCCGTACAAAACGGC.....AGTCT 2516
1168 AsnGlyArgAsnAspThrGlyThrValIleAsnLysAspGlyLeuThrIl 1184
2517 GACGCTTTCCGGCAACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACG 2566
1184 eThrLeuAlaAsnGlyAlaAlaAlaGlyThr.....AspAlaSerAsnG 1199
2567 GTAAT.....GTCTCCCTAGCCGATAAGGCA 2592
1199 lyAsnThrIleSerValThrLysAspGlyIleSerAlaGlyAsnLysGlu 1215
2593 GTATTCCATTTTGAAAGCAGCCGCTTTACCGGACAAATCAGCGGCGGCAA 2642
1216 IleThrAsnValLysSerAlaLeuLysThr.....TyrLysAspThrGl 1230
2643 GGATACGGCATTACACTTAAAAGACAGCGAATGG.....ACGCTGCCGT 2686
1230 nAsnThrAlaAspGluThrGlnAspLysGluPheHisAlaAlaValLysA 1247
2687 CAGGCACGGAATTAGGCAATTTAAACCTTGACAACGCCACCATTACA... 2733
1247 snAlaAsnGluValGluPheValGlyLysAsnGlyAlaThrValSerAla 1263
2734 ...CTCAATTCCGCCTATCGCCAC.....GATGCGGCAGGGGC 2768
1264 LysThrAspAsnAsnGlyLysHisThrValThrIleAspValAlaGluAl 1280
2769 GCAAACCGGCAGTGCG.....ACAGATGCG.....CCGCGCC 2800
1280 aLysValGlyAspGlyLeuGluLysAspThrAspGlyLysIleLysLeuL 1297
2801 GCCGTTCGCGCCGTTCGCGCCGTTCCCTATTATCCGTT.....ACACCG 2844
1297 ysValAspAsnThrAspGlyAsnAsnLeuLeuThrValAspAlaThrLys 1313
2845 CCAACTTCGGTA...GAATCCCGTTTCAACACGCTGACG..... 2880
1314 GlyAlaSerValAlaLysGlyGluPheAsnAlaValThrThrAspAlaTh 1330
2881 GTAAACGGCAAATTGAACGGTCAGGGAACATTCCGCT 2917

```
                  :::|||::::::   :::|||:::   :::   :::
1330 rThrAlaGlnGlyThrAsnAlaAsnGluArgGlyLysValValValLysG 1347

2918 TTATGTCGGAACTCTTCGGCTACCGCAGCGACAAATTGAAGCTGGCGGAA 2967
        :::      :::   ::::::||||||   |||:::|||
1347 lySerAsnGlyAlaThrAlaThrGluThrAspLysLysLysValAla... 1362

2968 AGTTCCGAAGGCACTTACACCTTGGCGGTCAACAATACCGGCAACGAACC 3017
        :::   |||      :::   |||:::|||::::::::::::
1363 ...ThrValGlyAspValAlaLysAlaIleAsnAspAlaAlaThr..... 1376

3018 TGCAAGCCTCGAACAATTGACGGTAGTGGAAGGAAAAGACAACAAACCGC 3067
                     ::::::   ||||||:::   |||:::      :
1377 ..............PheValLysValGluAsnAspAspSerAlaThrI 1388

3068 TGTCCGAAAACCTTAATTTCACCCTGCAAAACGAACACGTCGATGCCGGC 3117
     ::::::::::   :::            |||:::   :::   ||||||
1388 leAspAspSerProThrAspAspGlyAlaAsnAspAlaLeuLysAlaGly 1404

3118 ..................GCGTGGCGTTACCAACTCATCCGCAAAGACGG 3149
                        |||   :::         :::::::::|||||
1405 AspThrLeuThrLeuLysAlaGlyLysAsnLeuLysValLysArgAspGl 1421

3150 CGAG.........TTCCGCCTGCATAAT......CCGGTCAAAGAACAAG 3184
     |:::         |||   |||   |||         |||||:::
1421 yLysAsnIleThrPheAlaLeuAlaAsnAspLeuSerValLysSerAlaT 1438

3185 AGCTTTCCGACAAACTC..................GGCAAGGCAGAAGCC 3216
       :::||||||||||||                  :::|||::::::
1438 hrValSerAspLysLeuSerLeuGlyThrAsnGlyAsnLysValAsnIle 1454

3217 AAAAAACAGGCGGAA..............AAAGACAACGCGCAAAGCCT 3251
        :::::::::::             |||||||:::      :::
1455 ThrSerAspThrLysGlyLeuAsnPheAlaLysAspSerLysThrGlyAs 1471

3252 TGACGCGCTGATTGCGGCCGGGCGCGATGCCGTCGAAAAGACAGAAAGCG 3301
      ||||||   |||      :::      |||         |||::::::::
1471 pAspAlaAsnIleHisLeuAsnGlyIleAlaSerThrLeuThrAspThrL 1488

3302 TT......GCCGAACCGGCCCGGCAGGCAGGCGGGGAAAATGTCGGCATT 3345
     ::      :::      :::   :::   ||||||:::::::::
1488 euLeuAsnSerGlyAlaThrThrAsnLeuGlyGlyAsnGlyIle..... 1502

3346 ATGCAGGCGGAGGAAGAGAAAAAACGGGTGCAGGCGGATAAAGACACCGC 3395
        :::::::::|||||||||||||:::   :::   ||||||:::
1503 ......ThrAspAsnGluLysLysArgAlaAlaSerValLysAspValLe 1517

3396 CTTGGCGAAACAGCGCGAAGCGGAAACCCGGCCGGCTACCACCGCCTTCC 3445
       |||      :::         ::::::||||||:::
1517 uAsnAlaGlyTrpAsnValArgGlyValLysProAlaSer.......... 1530

3446 CCCGCGCCCGCCGCGCCCGCCGGGATTTGCCGCAACTGCAACCCCAACCG 3495

1530 ................................................. 1530

3496 CAGCCCCAACCGCAGCGCGACCTGATCAGCCGTTATGCCAATAGCGGTTT 3545
                                          ||||||:::   ::
1531 ..................................AlaAsnAsnGlnVa 1535

3546 GAGT......GAATTTTCCGCCACGCTCAACAGCGTTTTCGCCGTACAGG 3589
     ::::      :::|||   ||||||   ::::::|||      |||:::
1535 lGluAsnIleAspPheValAlaThrTyrAspThrValAspPheValSerG 1552

3590 ACGAATTAGAC......CGCGTATTTGCCGAAGACCGCCGCAAC...... 3627
       :::   |||         |||   :::|||::::::   |||
1552 lyAspLysAspThrThrSerValThrValGluSerLysAspAsnGlyLys 1568

3628 ..................GCCGTTTGGACAAGCGGCATCCGGGACACCAA 3659
                        ::::::   ||||||   |||:::|||
1569 ArgThrGluValLysIleGlyAlaLysThrSerValIleLysAsp..... 1583

3660 ACACTACCGTTCGCAAGATTTCCGCGCCTACCGCCAACAAACCGACCTGC 3709
        ::::::::::::   |||   :::   ::::::      |||   :
```

```
     1584 ....HisAsnGlyLysLeuPheThrGly...LysGluLeuLysAspAlaA 1598

     3710 GCCAAATCGGTATGCAG.................AAAAACCTCGGCAGC 3741
          :::::   |||:::                    |||:::   |||:::
     1598 snAsnAsnGlyValThrValThrGluThrAspGlyLysAspGluGlyAsn 1614

     3742 GGGCGCGTC.........GGCATCCTGTTTTCGCACAACCGG........ 3774
          |||   |||         :::::::::   :::   |||:::
     1615 GlyLeuValThrAlaLysAlaValIleAspAlaValAsnLysAlaGlyTr 1631

     3774 ..................................................... 3774

     1631 pArgValLysThrThrGlyAlaAsnGlyGlnAsnAspAspPheAlaThrV 1648

     3775 ...........ACCGAAAACACCTTCGACGACGGCATCGGCAACTCGGCA 3813
                      |||:::   ||||||   ||||||   |||::::::|||
     1648 alAlaSerGlyThrAsnValThrPheAlaAspGlyAsnGlyThrThrAla 1664

     3814 CGGCTTGCC.........CACGGCGCCGTTTTCGGGCAATACGGCATCGA 3854
          :::::::::         |||:::::::      :::|||::::::
     1665 GluValThrLysAlaAsnAspGlySerIleThrValLysTyrAsnVal.. 1680

     3855 CAGGTTCTACATCGGCATCAGCGCGGGCGCGGGTTTTAGCAGCGGCAGCC 3904
                                                          ::::
     1681 ..................................................LysV 1682

     3905 TTTCAGACGGCATC.........GGAGGCAAAATC 3930
          :::::||||||::         |||   ||||||
     1682 alAlaAspGlyLeuLysLeuAspGlyAspLysIle 1693

seq_name: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-268-347-47

seq_documentation_block:
; Sequence 47, Application US/09268347
; Patent No. 6335182
; GENERAL INFORMATION:
;  APPLICANT: Loosmore, Sheena M.
;  TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS
;  FILE REFERENCE: 1038-860
;  CURRENT APPLICATION NUMBER: US/09/268,347
;  CURRENT FILING DATE: 1999-03-16
;  NUMBER OF SEQ ID NOS: 54
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 47
;   LENGTH: 2354
;   TYPE: PRT
;   ORGANISM: Haemophilus influenzae
US-09-268-347-47


alignment_scores:
            Quality:  229.50            Length:   1495
              Ratio:  0.339               Gaps:     72
 Percent Similarity:  45.284   Percent Identity: 19.130

alignment_block:
 US-09-303-518D-649 x US-09-268-347-47   ..

 Align seg 1/1  to: US-09-268-347-47  from: 1  to: 2354

     172 GCCGAAAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTA 221
         ||:::::::   |||   :::         :::||||||   ||||||
     436 AlaAspThrAspGlyAlaLeuGluGlyIleSerLysAspGlnGluVal.. 451

     222 CAACAAAAAAGGGGAGTTGGTC............GGCAAATCAATG...A 256
             |||   ||||||   |||           ||||||::::::   :
     452 ....LysAlaGlyGluThrValThrPheLysAlaGlyLysAsnLeuLysV 467

     257 CAAAAGCCCCGATGATTGATTTTTCTGTGGTGTCGCGTAACGGCGTGGCG 306
         ::|||           :::|||:::         ::::::::::::::
     467 alLysGlnAspGlyAlaAsnPheThrTyrSerLeuGlnAspAlaLeuThr 483

     307 GCATTGGTGGGC..................................... 318
```

```
    :::|||::::::
484 GlyLeuThrSerIleThrLeuGlyGlyThrThrAsnGlyGlyAsnAspAl 500

319 ..................GATCAATATATTGTGAGCGTGGCACATAACG 349
                       |||         ::::::   |||   ||||
500 aLysThrValIleAsnLysAspGlyLeuThrIleThrProAlaGlyAsnG 517

350 GCGGCTATAACAACGTTGATTTTGGTGCGGAAGGAAGAAATCCCGATCAA 399
    ||                  |||:::   |||   |||
517 ly....................GlyThrThrGlyThrAsn......... 523

400 CATCGTTTTACTTATAAAATTGTGAAACGGAATAATTATAAAGCAGGGAC 449
          |||   ::::::::::|||   ::::::   |||||||||::
524 .........ThrIleSerValThrLys...AspGlyIleLysAlaGlyAs 536

450 TAAAGGCCATCCTTATGGCGGCGATTATCATATGCCGCGTTTGCATAAAT 499
    ::|||:::
536 nLysAla.......................................... 538

500 TTGTCACAGATGCAGAACCTGTTGAAATGACCAGTTATATGGATGGGCGG 549
                              :::|||:::      :::|||
539 .........................IleThrAsnValAlaSerGlyLeu 546

550 AAATATATCGATCAAAATAATTACCCTGACCGTGTTCGTATTGGGGCAGG 599
    :::      |||:::   |||:::
547 ArgAlaTyrAspAspAlaAsnPhe........................... 554

600 CAGGCAATATTGGCGATCTGATGAAGATGAGCCCAATAAC......CGCG 643
                            |||      ||||||         :
555 .........................AspValLeuAsnAsnSerAlaThrA 563

644 AAAGTTCATATCATATTGCAAGTGCGTATTCTTGGCTCGTTGGTGGCAAT 693
    ::   :::   |||:::   :::|||||||:::
563 spLeuAsnArgHisValGluAspAlaTyrLys.................. 573

694 ACCTTTGCACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGTGA 743
                             |||   :::||||||:::::::::
574 ............................GlyLeuLeuAsnLeuAsnGluLy 581

744 AAAAATTAAACATAGCCCATATGGTTTTTTACCAACAGGAGGCTCATTTG 793
    ::::   :::   :::|||               |||:::::::::   |
581 sAsnAlaAsnLysGlnProLeuValThrAspSerThrAlaAlaThrValG 598

794 GCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGG 843
    ||                              |||   :::|||      |||
598 ly...........................AspLeuArgLysLeuGlyTrp 605

844 TTAATT.........AATGGGGTATTGCAAACGGGCAACCCCTATATAGG 884
    ::::::        ||||||:::   :::   :::|||
606 ValValSerThrLysAsnGlyThrLysGluGluSerAsn........... 618

885 AAAAAGCAATGGCTTCCAGCTGGTTCGTAAAGATTGGTTCTATGATGAAA 934
                  |||:::   :::   |||         :::   :
619 ................GlnValLysGlnAlaAsp.........GluValL 627

935 TCTTTGCTGGAGATACCCATTCAGTATTCTACGAACCACGTCAAAATGGG 984
    ::|||:::|||         ::::::      :::      :::||||||
627 euPheThrGlyAlaGlyAlaAlaThrValThrSerLysSerGluAsnGly 643

985 AAATACTCTTTTAACGACGATAATAATGGCACAGGAAAAATCAATGCCAA 1034
    |||::::::
644 LysHisThr....................................... 646

1035 ACATGAACACAATTCTCTGCCTAATAGATTAAAAACACGAACCGTTCAAT 1084

646 ............................................... 646

1085 TGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTGTTTATCATGCT 1134
       ::::::||||||::::::||||||      :::         :::
647 ..IleThrValSerValAlaGluThrLysAlaAspCysGlyLeuGluLys 662

1135 GCAGGTGGTGTCAACAGTTATCGACCCAGACTGAATAATGGAGAAAATAT 1184
       |||   :::   :::   :::         :::|||   :::||||::

663 AspGlyAspThrIleLysLeuLysValAspAsnGlnAsnThrAspAsnVa 679

1185 TTCCTTTATT..................................... 1194
     :      :::
679 lLeuThrValGlyAsnAsnGlyThrAlaValThrLysGlyGlyPheGluT 696

1195 ...................GACGAAGGAAAAGGCGAATTGATACTT... 1221
                         |||      :::|||::::::   :::
696 hrValLysThrGlyAlaThrAspAlaAspArgGlyLysValThrValLys 712

1222 ......ACCAGCAAC.............................. 1230
           |||:::|||
713 AspAlaThrAlaAsnAspAlaAspLysLysValAlaThrValLysAspVa 729

1231 ..........ATCAATCAAGGTGCTGGAGGATTATATTTCCAAGGAGATT 1270
               ||||||::::::|||      :::      ::::::
729 lAlaThrAlaIleAsnSerAlaAlaThrPheValLysThrGluAsnLeuT 746

1271 TTACGGTCTCGCCTGAAAATAACGAAACTTGGCAAGGCGCGGGCGTTCAT 1320
       |||         |||:::|||   |||   :::|||
746 hrThrSerIleAspGluAspAsnProThrAspAsnGlyLysAspAspAla 762

1321 ATCAGTGAAGACAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGA 1370
     ::::::      :::|||:::|||:::|||::::::          |||
763 LeuLysAlaGlyAspThrLeuThrPheLysAlaGly......LysAsnLe 777

1371 CCGCCTGTCCAAAATCGGCAAAGGC...................... 1395
      ::::::::::::   ||||||:::
777 uLysValLysArgAspGlyLysAsnIleThrPheAspLeuAlaLysAsnL 794

1396 ............................ACGCTGCACGTTCAAGCCAAA 1416
                                 ||||||   :::   ::::::
794 euGluValLysThrAlaLysValSerAspThrLeuThrIleGlyGlyAsn 810

1417 GGGGAAAACCAAGGCTCG.......................ATCAGCGT 1442
           :::   |||:::                            |||::
811 ThrProThrGlyGlyThrThrAlaThrProLysValAsnIleThrSerTh 827

1443 GGGCGACGGTACAGTCATTTTGGATCAGCAGGCAGACGATAAAGGCAAAA 1492
     ::::||||||      :::      :::   ||||||   :::|||:::|
827 rAlaAspGlyLeuAsnPheAlaLysGluThrAlaAspAlaSerGlySerL 844

1493 AACAAGCCTTT............................AGTGAAATC 1512
     ||:::::::::                            :::
844 ysAsnValTyrLeuLysGlyIleAlaThrThrLeuThrGluProSerAla 860

1513 GGCTTGGTCAGCGGCAGGGGTACGGTGCAACTGAATGCCGATAATCAGTT 1562
     |||      |||::::::      :::::::::::|||   ::::::
861 GlyAlaLysSerSerHisValAspLeuAsnValAspAlaThrLysLysSe 877

1563 CAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGA...CGTTTGGATT 1609
       |||      ::::::      :::|||:::|||   :::::::::
877 rAsnAlaAlaSerIleGluAspValLeuArgAlaGlyTrpAsnIleGlnG 894

1610 TAAACGGGCATTCGCTTTCGTTCCACCGTATTCAAAATACCGATGAAGGG 1659
       ||||||:::::::::::::::            :::|||
894 lyAsnGlyAsnAsnValAspTyrValAlaThrTyrAspThr......... 907

1660 GCGATGATTGTCAACCACAATCAAGACAAAGAATCCACCGTTACCATTAC 1709
               ||||||   :::::::|||:::   :::|||:::|||:::||
908 .........ValAsnPheThrAspAspSerThrGlyThrThrThrValTh 921

1710 AGGCAATAAAGATATTGCTACAACCGGCAATAACAACAGCTTGGATAGCA 1759
     |   ::::::                |||   :::   ::::::   :::|
921 rValThrGlnLysAlaAspGlyLysGlyAlaAspValLysIleGlyAlaL 938

1760 AAAAAGAAATTGCC......TACAACGGT...TGGTTTGGCGAGAAAGAT 1800
     ||   ::::::         :::||||||      |||      ||||||
938 ysThrSerValIleLysAspHisAsnGlyLysLeuPheThrGlyLysAsp 954

1801 ACGACCAAAACGAACGGGCGGCTCAACCTTGTTTACCAGCCCGCCGCAGA 1850
               :::|||:::      ::::::      :::   :::   ::
955 LeuLysAspAlaAsnAsnGlyAlaThrValSerGluAspAspGlyLysAs 971
```

```
1851 AGACCGCACCCTGCTGCTTTCCGGCGGAACA.............AATT 1885
     :      |||   |||:::::::::   |||               |||
 971 pThrGlyThrGlyLeuValThrAlaLysThrValIleAspAlaValAsnL 988

1886 TAAACGGCAACATCACGCAAACAAACGGCAAACTGTTTTTCAGCGGCAGA 1935
       :::|||      :::      :::|||           :::|||
 988 ysSerGlyTrpArgValThrGlyGluGlyAlaThrAlaGluThrGlyAla 1004

1936 CCAACACCGCACGCCTACAATCATTTAAACGACCATTGGTCGCAAAAAGA 1985
       :::   :::|||  |||                          ||
1005 ThrAlaValAsnAlaGlyAsn.....................AlaGl 1013

1986 GGGCATTCCTCGCGGGGAAATCGTGTGGGACAACGACTGGATCAACCGCA 2035
     |   :::      |||     |||
1013 uThrValThrSerGlyThrSerVal..................... 1021

2036 CATTTAAAGCGGAAAACTTCCAAATTAAAGGCGGACAGGCGGTGGTTTCC 2085
                 ||||||     |||:::|||:::|||:::::::::
1022 .............AsnPhe......LysAsnGlyAsnAlaThrThrAla 1031

2086 CGCAATGTTGCCAAAGTGAAAGGCGATTGGCATTTGAGCAATCACGCCCA 2135
       :::|||:::|||   :::|||:::   :::::::::      :::::
1032 ...ThrValSerLysAspAsnGlyAsnIleAsnValLysTyrAspValAs 1047

2136 AGCAGTTTTTGGTGTCGCACCGCATCAAAGCCACACAATCTGTACACGTT 2185
     ::::      |||:::         ::::::      |||        :
1047 nValGlyAspGlyLeuLysIleGlyAspAspLysLysIleVal......A 1062

2186 CGGACTGGACGGGTCTGACAAATTGTGTCGAAAAAACCATTACCGACGAT 2235
     ::|||   |||   |||                  |||:::|||
1062 laAspThrThrThrLeu.................ThrValThrGlyGly 1072

2236 AAAGTGATTGCTTCATTGACTAAGACCGACATCAGCGGCAATGTCGATCT 2285
     ||||||   :::              :::::::::::::|||      ||
1073 LysValSerValProAlaGlyAlaAsnSerValAsnAsnAsnLysLysLe 1089

2286 TGCCGATCACGCTCATTTAAATCTCACAGGGCTTGCCACACTCAACGGCA 2335
     |               :::|||      ||||||||||||      :::|
1089 u..............ValAsnAlaGluGlyLeuAlaThrAlaLeuAsnA 1101

2336 ATCTTAGT.......................GCAAATGGCGATACACGT 2361
     ||||||||                      |||:::|||:::::::::
1101 snLeuSerTrpThrAlaLysAlaAspLysTyrAlaAspGlyGluSerGlu 1117

2362 TATACAGTCAGCCACAACGCCACCCAAAACGGCAACCTTAGCCTC..... 2406
           :::::::::::::::      :::   ::::::::::::
1118 GlyGluThrAspGlnGluValLysAlaGlyAspLysValThrPheLysAl 1134

2407 ................GTGGGCAATGCCCAAGCAACATTTAAT....... 2433
                     |||   :::::::::      |||:::
1134 aGlyLysAsnLeuLysValLysGlnSerGluLysAspPheThrTyrSerL 1151

2434 ..CAAGCCACATTAAACGGCAACACATCGGCTTCG......GGCAATGCT 2475
       |||   ||||||:::|||   ||||||   :::      |||:::|||
1151 euGlnAspThrLeuThrGlyLeuThrSerIleThrLeuGlyGlyThrAla 1167

2476 TCATTTAATCTAAGCGACCACGCCGTACAAAACGGC.........AGTCT 2516
     :::   :::   :::      :::|||   |||             :::::
1168 AsnGlyArgAsnAspThrGlyThrValIleAsnLysAspGlyLeuThrIl 1184

2517 GACGCTTTCCGGCAACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACG 2566
     :||||||:::::::::|||   |||:::::      :::|||   ||||
1184 eThrLeuAlaAsnGlyAlaAlaAlaGlyThr......AspAlaSerAsnG 1199

2567 GTAAT.......................GTCTCCCTAGCCGATAAGGCA 2592
     |||||                      :::|||   :::::::|||
1199 lyAsnThrIleSerValThrLysAspGlyIleSerAlaGlyAsnLysGlu 1215

2593 GTATTCCATTTTGAAAGCAGCCGCTTTACCGGACAAATCAGCGGCGGCAA 2642
     :::   :::   :::|||:::      |||      :::      ::
1216 IleThrAsnValLysSerAlaLeuLysThr......TyrLysAspThrGl 1230
```

```
2643 GGATACGGCATTACACTTAAAAGACAGCGAATGG......ACGCTGCCGT 2686
     ::::||||||   :::   :::|||:::|||:::      ::::::   :
1230 nAsnThrAlaAspGluThrGlnAspLysGluPheHisAlaAlaValLysA 1247

2687 CAGGCACGGAATTAGGCAATTTAAACCTTGACAACGCCACCATTACA... 2733
     ::::::::|||:::      ::::::   ::::::||||||::::::
1247 snAlaAsnGluValGluPheValGlyLysAsnGlyAlaThrValSerAla 1263

2734 ...CTCAATTCCGCCTATCGCCAC............GATGCGGCAGGGGC 2768
          ::::::      :::|||              |||:::|||   ||
1264 LysThrAspAsnAsnGlyLysHisThrValThrIleAspValAlaGluAl 1280

2769 GCAAACCGGCAGTGCG............ACAGATGCG......CCGCGCC 2800
     |:::::::|||::::::           ||||||:::           :
1280 aLysValGlyAspGlyLeuGluLysAspThrAspGlyLysIleLysLeuL 1297

2801 GCCGTTCGCGCCGTTCGCGCCGTTCCCTATTATCCGTT......ACACCG 2844
     ::   ::::::   :::   ::::::||||||:::|||      |||
1297 ysValAspAsnThrAspGlyAsnAsnLeuLeuThrValAspAlaThrLys 1313

2845 CCAACTTCGGTA...GAATCCCGTTTCAACACGCTGACG........... 2880
        :::||||||   :::::::::||||||::::::|||
1314 GlyAlaSerValAlaLysGlyGluPheAsnAlaValThrThrAspAlaTh 1330

2881 .............GTAAACGGCAAATTGAACGGTCAGGGAACATTCCGCT 2917
                    :::|||::::::   :::|||:::   :::   :::
1330 rThrAlaGlnGlyThrAsnAlaAsnGluArgGlyLysValValValLysG 1347

2918 TTATGTCGGAACTCTTCGGCTACCGCAGCGACAAATTGAAGCTGGCGGAA 2967
          :::          :::   ::::::||||||   |||:::|||
1347 lySerAsnGlyAlaThrAlaThrGluThrAspLysLysLysValAla... 1362

2968 AGTTCCGAAGGCACTTACACCTTGGCGGTCAACAATACCGGCAACGAACC 3017
        :::   |||       :::   |||:::|||::::::::::::
1363 ...ThrValGlyAspValAlaLysAlaIleAsnAspAlaAlaThr..... 1376

3018 TGCAAGCCTCGAACAATTGACGGTAGTGGAAGGAAAAGACAACAAACCGC 3067
                       ::::::   ||||||:::   |||:::      :
1377 ................PheValLysValGluAsnAspAspSerAlaThrI 1388

3068 TGTCCGAAAACCTTAATTTCACCCTGCAAAACGAACACGTCGATGCCGGC 3117
     ::::::::::::   :::             |||:::   :::   ||||||
1388 leAspAspSerProThrAspAspGlyAlaAsnAspAlaLeuLysAlaGly 1404

3118 ..................GCGTGGCGTTACCAACTCATCCGCAAAGACGG 3149
                        |||   :::          :::::::::||||
1405 AspThrLeuThrLeuLysAlaGlyLysAsnLeuLysValLysArgAspGl 1421

3150 CGAG.........TTCCGCCTGCATAAT......CCGGTCAAAGAACAAG 3184
     |:::         |||   |||   |||          ||||||:::
1421 yLysAsnIleThrPheAlaLeuAlaAsnAspLeuSerValLysSerAlaT 1438

3185 AGCTTTCCGACAAACTC..................GGCAAGGCAGAAGCC 3216
       :::||||||||||||                  :::|||::::::
1438 hrValSerAspLysLeuSerLeuGlyThrAsnGlyAsnLysValAsnIle 1454

3217 AAAAAACAGGCGGAA..............AAAGACAACGCGCAAAGCCT 3251
        ::::::::::                |||||||:::      :::
1455 ThrSerAspThrLysGlyLeuLysPheAlaLysAspSerLysThrGlyAs 1471

3252 TGACGCGCTGATTGCGGCCGGGCGCGATGCCGTCGAAAAGACAGAAAGCG 3301
      ||||||   |||      :::      |||         |||::::::::
1471 pAspAlaAsnIleHisLeuAsnGlyIleAlaSerThrLeuThrAspThrL 1488

3302 TT......GCCGAACCGGCCCGGCAGGCAGGCGGGGAAAATGTCGGCATT 3345
     ::      :::      :::   :::   ||||||::::::::::
1488 euLeuAsnSerGlyAlaThrThrAsnLeuGlyGlyAsnGlyIle...... 1502

3346 ATGCAGGCGGAGGAAGAGAAAAAACGGGTGCAGGCGGATAAAGACACCGC 3395
           :::::::::|||||||||||:::   :::   |||||||:::
1503 ......ThrAspAsnGluLysLysArgAlaAlaSerValLysAspValLe 1517

3396 CTTGGCGAAACAGCGCGAAGCGGAAACCCGGCCGGCTACCACCGCCTTCC 3445
```

```
          |||      :::         ::::::||||||:::
    1517 uAsnAlaGlyTrpAsnValArgGlyValLysProAlaSer..... 1530

    3446 CCCGCGCCCGCCGCGCCCGCCGGGATTTGCCGCAACTGCAACCCCAACCG 3495

    1530 ..................................................... 1530

    3496 CAGCCCCAACCGCAGCGCGACCTGATCAGCCGTTATGCCAATAGCGGTTT 3545
                                            ||||||:::   ::
    1531 .................................AlaAsnAsnGlnVa 1535

    3546 GAGT......GAATTTTCCGCCACGCTCAACAGCGTTTTCGCCGTACAGG 3589
         ::::      :::|||   ||||||   ::::::|||      |||:::
    1535 lGluAsnIleAspPheValAlaThrTyrAspThrValAspPheValSerG 1552

    3590 ACGAATTAGAC......CGCGTATTTGCCGAAGACCGCCGCAAC...... 3627
           :::   |||         |||   :::|||::::::   |||
    1552 lyAspLysAspThrThrSerValThrValGluSerLysAspAsnGlyLys 1568

    3628 ..................GCCGTTTGGACAAGCGGCATCCGGGACACCAA 3659
                           ::::::   ||||||   |||:::|||
    1569 ArgThrGluValLysIleGlyAlaLysThrSerValIleLysAsp..... 1583

    3660 ACACTACCGTTCGCAAGATTTCCGCGCCTACCGCCAACAAACCGACCTGC 3709
         ::::::::::::   |||   :::   ::::::      |||   :
    1584 ....HisAsnGlyLysLeuPheThrGly...LysGluLeuLysAspAlaA 1598

    3710 GCCAAATCGGTATGCAG..................AAAAACCTCGGCAGC 3741
         :::::   |||:::                     |||:::   |||:::
    1598 snAsnAsnGlyValThrValThrGluThrAspGlyLysAspGluGlyAsn 1614

    3742 GGGCGCGTC.........GGCATCCTGTTTTCGCACAACCGG........ 3774
         |||   |||         :::::::::   :::   |||:::
    1615 GlyLeuValThrAlaLysAlaValIleAspAlaValAsnLysAlaGlyTr 1631

    3774 ..................................................... 3774

    1631 pArgValLysThrThrGlyAlaAsnGlyGlnAsnAspAspPheAlaThrV 1648

    3775 ..........ACCGAAAACACCTTCGACGACGGCATCGGCAACTCGGCA 3813
                   |||:::   ||||||   ||||||   |||::::::|||
    1648 alAlaSerGlyThrAsnValThrPheAlaAspGlyAsnGlyThrThrAla 1664

    3814 CGGCTTGCC.........CACGGCGCCGTTTTCGGGCAATACGGCATCGA 3854
         :::::::::         |||::::::      :::|||::::::
    1665 GluValThrLysAlaAsnAspGlySerIleThrValLysTyrAsnVal.. 1680

    3855 CAGGTTCTACATCGGCATCAGCGCGGGCGCGGGTTTTAGCAGCGGCAGCC 3904
                                                    ::::
    1681 ..................................................LysV 1682

    3905 TTTCAGACGGCATC.........GGAGGCAAAATC 3930
         :::::||||||:::         |||   ||||||
    1682 alAlaAspGlyLeuLysLeuAspGlyAspLysIle 1693

seq_name: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-968-685A-10

seq_documentation_block:
; Sequence 10, Application US/08968685A
; Patent No. 6214981
;  GENERAL INFORMATION:
;    APPLICANT:  TUCKER, KENNETH
;    APPLICANT:  PLOSILA, LAURA
;    TITLE OF INVENTION:  MORAXELLA CATARRHALIS OUTER MEMBRANE
;    TITLE OF INVENTION:  PROTEIN-106 POLYPEPTIDE, GENE SEQUENCE AND USES THEREOF
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PENNIE & EDMONDS LLP
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/968,685A
;      FILING DATE:  No. 6214981ember 12, 1997
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Baldwin, Geraldine F.
;      REGISTRATION NUMBER:  31,232
;      REFERENCE/DOCKET NUMBER:  7969-060
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  (212) 869-8864
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2123 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  peptide
US-08-968-685A-10


alignment_scores:
            Quality:  223.00              Length:   1548
              Ratio:  0.318                 Gaps:     71
 Percent Similarity:  45.349   Percent Identity: 18.992

alignment_block:
 US-09-303-518D-649 x US-08-968-685A-10   ..

 Align seg 1/1  to: US-08-968-685A-10  from: 1  to: 2123

     469 GGCGATTATCATATGCCGCGTTTGCATAAATTTGTCACAGATGCAGAACC 518
         ||||||      :::   :::   :::      :::|||   |||
     779 GlyAspLeuThrPheSerLysGlnAsnThrGlyAlaThrProAlaThrAs 795

     519 TGTTGAAATGACCAGTTATATGGATGGGCGGAAATATATCGATCAAAAT. 567
           :::             :::||||||   |||:::   |||:::|||
     795 nSerLysThrIleTyrGlyValAspGlyLeuLysPheThrAspAsnAsnG 812

     568 ....................AATTACCCTGACCGTGTTCGTATTGGGGCA 597
                             :::|||      :::   :::::|||
     812 lyIleAlaLeuAspGlyThrThrTyrIleThrLysAspLysValGlyPhe 828

     598 GGCAGGCAATATTGGCGATCTGATGAAGATGAGCCCAATAACCGCGAAAG 647
         ::::::|||            |||:::::::::|||         :::::
     829 AlaLysGlnAspGlySerLeuAspLysSerLysProTyrLeuAspLysAs 845

     648 TTCATATCATATT..................................... 660
         ::::     :::
     845 pLysLeuLysValGlyGluValGluIleThrThrAsnGlyIleAsnAlaG 862

     661 ..GCAAGTGCGTATTCTTGGCTCGTTGGTGGCAATACCTTTGCACAAAAT 708
           ::::::|||   :::   |||   :::      |||   |||   |||
     862 lyGlyLysAlaIleThrGlyLeuSerAsnThrLeuThrAspAlaThrAsn 878

     709 GGATCAGGTGGTGGCACAGTCAACTTAGGT.....AGTGAAAA 746
         ::::::   |||   :::::::::||||||          ::::::||
     879 AlaThrThrGlyHisValThrGlnLeuGlyIleValAspSerThrAspLy 895

     747 AATTAAACATAGCCCATATGGTTTTTTACCAACAGGAGGC.......... 786
         |   :::   :::      |||   :::   ::::::|||
     895 sThrArgAlaAlaSerIleGlyAspValLeuAsnAlaGlyPheAsnLeuL 912

     787 ..TCATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAG 834
           :::   ||||||:::   :::   :::      ||||||:::
     912 ysAsnAsnGlyAspAlaLysAspPheValSerThrTyrAspThrValAsp 928
```

```
 835 CAAAAGTGGTTAATTAATGGGGTATTGCAAACGGGCAAC...CCCTATAT 881
             :::|||||||||        |||::::::     |||
 929 .........PheIleAsnGlyAsnAlaThrThrAlaLysValThrTyrAs 942

 882 AGGAAAAAGCAATGGCTTCCAGCTGGTTCGTAAAGATTGGTTCTATGATG 931
      ||||||::::::                                ||||||
 942 pGlyLysAlaSerLysValAla.....................TyrAspV 952

 932 AAATCTTTGCTGGAGATACC...CATTCAGTATTCTACGAACCACGTCAA 978
               |||   |||   |||   :::      :::   ::::::
 952 alAsnValAspGlyThrThrIleHisLeuThrGlyAlaAspGlyAsnLys 968

 979 AAT.............................................GG 983
     |||                                             ||
 969 AsnGlnIleGlyValLysThrThrThrLeuThrLysThrAspAlaLysGl 985

 984 GAAATACTCTTTTAAC...GACGATAATAATGGCACAGGAAAA......A 1024
     |      ::::::|||   :::   |||:::|||      |||      |
 985 yAspLysAlaIleAsnPheSerValAsnSerGlyAspAspLysAlaLeuI 1002

1025 TCAATGCCAAACATGAACACAATTCTCTGCCTAATAGATTAAAAACACGA 1074
     ||||||||||           ::::::   ::::::|||:::|||
1002 leAsnAlaLys...........AspIleAlaAspAsnLeuAsnThrLeu 1014

1075 ACCGTTCAATTGTTTAATGTT...TCTTTATCCGAGACAGCAAGAGAACC 1121
     :::   ::::::   |||:::   :::   ::::::||||||   :::
1015 AlaGlyGluIleArgAsnThrLysGlyThrAlaAspThrAlaLeuGlnTh 1031

1122 TGTTTATCATGCTGCAGGTGGTGTCAACAGTTATCGACCCAGACTGAATA 1171
                          |||::::::   :::   :::   ::::
1031 rPheGln..............ValLysLysValLysGluAsnGlyAspA 1043

1172 ATGGAGAAAATATTTCCTTTATTGACGAAGGAAAAGGCGAATTGATACTT 1221
     ::   ::::::   :::   |||      ||||||
1043 spAspAsnAspAlaAspThrIleThrValGlyLysAsp......AlaLys 1057

1222 ACCAGCAACATCAATCAAGGTGCTGGAGGATTATATTTCCAAGGAGATTT 1271
     |||::::::::::|||               |||   ::::::|||'
1058 ThrAsnGlnValAsnThr............LeuLysLeuLysGlyLys.. 1069

1272 TACGGTCTCGCCTGAAAATAACGAAACTTGGCAAGGCGCGGGCGTTCATA 1321
                                   :::|||      ::::::
1070 ............................AsnGlyLeuAspIleGlnT 1076

1322 TCAGTGAAGACAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGAC 1371
       ::::::|||:::|||||||||:::   :::|||
1076 hrAsnLysAspGlyThrValThrPheGlyIleAsn.............. 1087

1372 CGCCTGTCCAAAATCGGCAAAGGCACGCTGCACGTTCAAGCCAAAGGGGA 1421

1087 ..................................................1087

1422 AAACCAAGGCTCGATCAGCGTGGGCGACGGTACAGTCATTTTGGATCAGC 1471
      :::|||:::::::::::::::|||::::::|||::::::   ::::::
1088 .ThrGlnSerGlyLeuLysAlaGlyAsnAsnThrThrLeuAsnAsnAsnG 1104

1472 AGGCAGACGATAAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCTTGGTC 1521
          :::   |||:::                          :::
1104 lyLeuSerIleLysAsn............................Thr 1110

1522 AGCGGCAGGGGTACGGTGCAACTGAATGCCGATAATCAGTTCAACCCCGA 1571
     :::|||:::      :::|||::::::||||||
1111 AlaGlyAsnGluGlnIleGlnValGlyAlaAsp.............. 1121

1572 CAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAACGGGCATT 1621
                                                    :
1122 ..................................................G 1122

1622 CGCTTTCGTTCCACCGTATTCAAAATACCGATGAAGGGGCGATGATTGTC 1671
     ::::::::|||   :::::::::|||        ||||||   |||
1122 lyValLysPheAlaLysValAsnAsnGlyValValGlyAlaGlyIle... 1137

1672 AACCACAATCAAGACAAAGAATCCACCGTTACCATTACAGGCAATAAAGA 1721
```

```
                        ::::::|||:::   ||||||   :::   ::
1138 ..................AspGlyThrThrArgIleThr...ArgAspGl 1147

1722 TATTGCTACAACCGGCAATAACAACAGCTTGGAT...AGCAAAAAAGAAA 1768
     :|||:::   :::|||:::|||:::||||||||   ||||||   ::::
1147 uIleGlyPheAlaGlyThrAsnGlySerLeuAspLysSerLysProHisL 1164

1769 TTGCCTACAACGGT......TGGTTTGGCGAGAAAGATACGACCAAAACG 1812
     :::::   :::|||           |||:::|||   |||:::
1164 euSerLysAspGlyIleAsnAlaGlyGlyLysLysIleThrAsnIleGln 1180

1813 AACGGGCGGCTCAACCTTGTTTACCAGCCCGCCGCAGAAGACCGCACCCT 1862
     :::|||::::::                     |||::::::   :::
1181 SerGlyGluIle.....................AlaGlnAsnSerAsnAs 1190

1863 GCTGCTTTCCGGCGGA.................ACAAATTTAAACGGCA 1894
         ::::::||||||                  |||:::|||:::::::
1190 pAlaValThrGlyGlyLysIleTyrAspLeuLysThrGluLeuGluAsnL 1207

1895 ACATCACGCAAACA..............AACGGCAAACTGTTTTTCAGC 1929
     ::|||::::::|||             |||:::         ||||||
1207 ysIleSerSerThrAlaLysThrAlaGlnAsnSerLeuHisGluPheSer 1223

1930 GGCAGACCAACACCGCACGCCTACAATCATTTAAACGACCATTGGTCGCA 1979
                                 :::::::::   :::|||::
1224 ValAlaAspGluGlnGlyAsnAsnPheThrValSerAsnProTyrSerSe 1240

1980 AAAAGAGGGCATTCCTCGCGGGGAAATCGTGTGGGACAACGACTGGATCA 2029
     :   :::            ::::::::::::
1240 rTyrAspThrSerLysThrSerAspValIle..................... 1250

2030 ACCGCACATTTAAAGCGGAAAAC...TTCCAAATTAAAGGCGGACAGGCG 2076
         ||||||   :::||||||   :::      |||   :::::::::
1251 .....ThrPheAlaGlyGluAsnGlyIleThrThrLysValAsnLysGly 1265

2077 GTGGTTTCCCGCAATGTTGCCAAAGTGAAAGGCGATTGGCATTTG..... 2121
     ||||||      ::::::   ::::::||||||         |||
1266 ValValArgValGlyIleAspGlnThrLysGly.........LeuThrTh 1279

2122 ..................AGCAATCACGCCCAAGCAGTTTTTGGTGTCG 2152
                         :::|||:::::::::::::::::        :
1279 rProLysLeuThrValGlyAsnAsnAsnGlyLysGlyIleValIleAspS 1296

2153 CACCGCATCAAAGCCACACAATCTGTACACGTTCGGACTGGACGGGTCTG 2202
     ::   :::   ::::::||||||                  |||||||||
1296 erGlnAsnGlyGlnAsnThrIle..................ThrGlyLeu 1306

2203 ACAAATTGTGTCGAA.....:...................AAAACCAT 2225
     :::|||   :::                            :::|||
1307 SerAsnThrLeuAlaAsnValThrAsnAspLysGlySerValArgThrTh 1323

2226 TACCGACGATAAAGTGATTGCTTCATTGACTAAGACCGACATCAGCGGCA 2275
         :::   |||:::|||   :::      ||||||      ::::::
1323 rGluGlnGlyLysIleIleLysAspGluAspLysThrArgAlaAlaSerI 1340

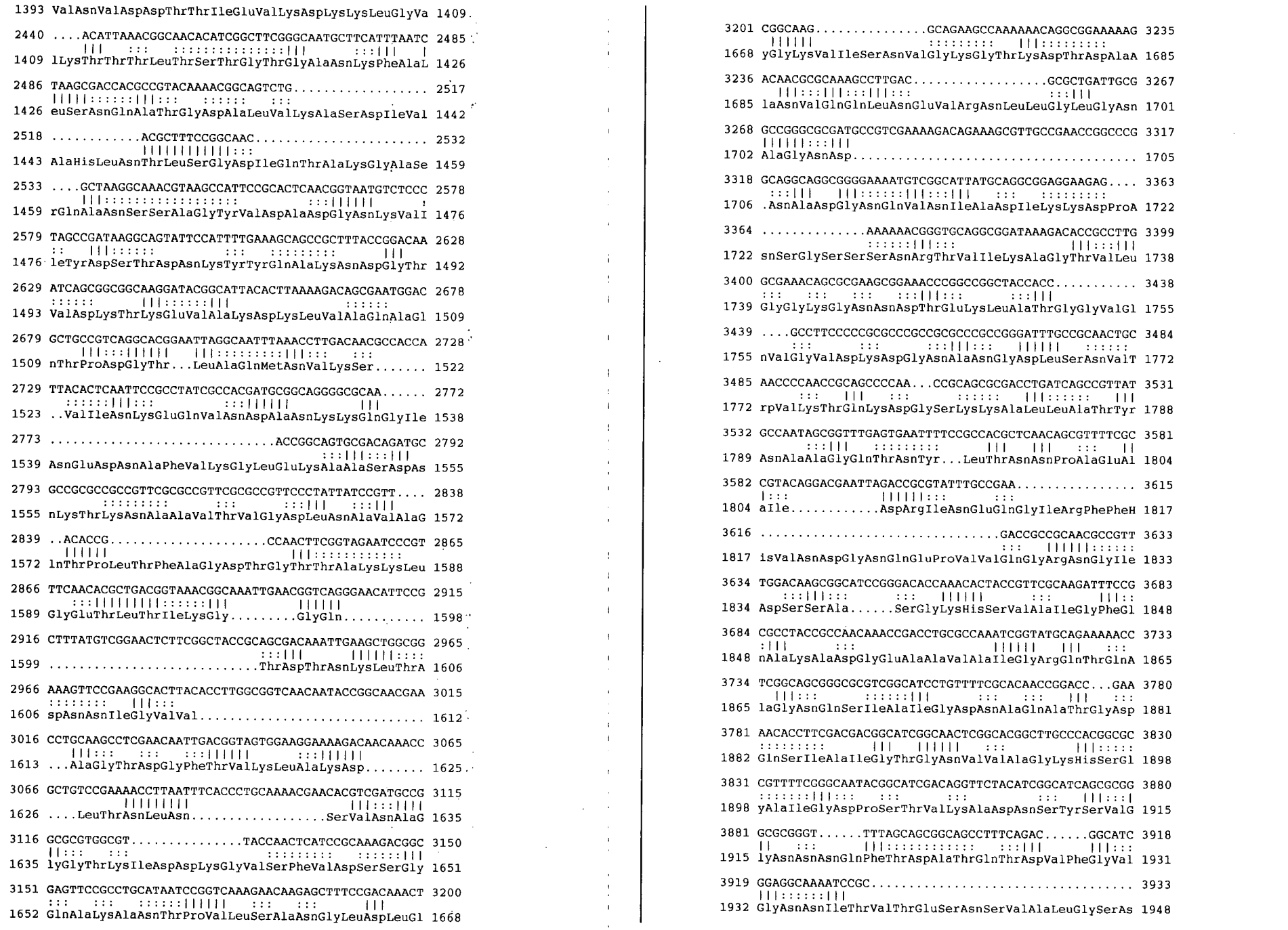
2276 ATGTCGATCTTGCCGATCACGCTCATTTAAATCTCACAGGGCTTGCCACA 2325
       ||||||                                  :::   :::
1340 leValAsp................................ValLeuSer 1345

2326 CTCAACGGCAATCTTAGTGCAAATGGCGATACACGTTATACAGTCAGCCA 2375
        :::   ||||||::::::||||||::::::         ||||||
1346 AlaGlyPheAsnLeuGlnGlyAsnGlyGluAlaValAspPheValSerTh 1362

2376 CAACGCCACCCAAAAC.........GGCAACCTTAGCCTCGTGGGCAATG 2416
           |||   |||         ||||||   :::          :::|
1362 rTyrAspThrValAsnPheAlaAspGlyAsnAlaThr.........ThrA 1376

2417 CCCAAGCAACATTTAATCAAGCC..........................  2439
     ||::::::|||::::::::::::
1376 laLysValThrTyrAspAspThrSerLysThrSerLysValValTyrAsp 1392

2439 ..................................................  2439
```



```
    3934 .....CGCCGCGTGCTGCATTACGGCATTCAGGCACGATACCGCG 3973
                              |||   |||   ||||||:::
    1948 nSerAlaIleSerAlaGlyThrHisAlaGlyThrGlnAlaLysLysSerA 1965

    3974 CCGGTTTCGGC..................................... 3984
           |||   :::
    1965 spGlyThrAlaGlyThrThrThrThrAlaGlyAlaThrGlyThrValLys 1981

    3985 GGATTCGGCATCGAACCGCACATCGGCGCAACGCGCTATTTCGTCCAAAA 4034
         ||||||:::   :::      :::|||||||::         ::::::
    1982 GlyPheAlaGlyGlnThrAlaValGlyAlaValSerValGlyAlaSerGl 1998

    4035 AGCGGATTACCGCTACGAAAACGTCAATATC.................. 4065
          |||:::   |||   :::||||||
    1998 yAlaGluArgArgIleGlnAsnValAlaAlaGlyGluValSerAlaThrS 2015

    4066 ................................GCCACCCCCGGCCTT 4080
                                        ||||||   |||:::
    2015 erThrAspAlaValAsnGlySerGlnLeuTyrLysAlaThrGlnGlyIle 2031

    4081 GCA.....................................TTCAACCG 4091
         |||                                    |||::
    2032 AlaAsnAlaThrAsnGluLeuAspHisArgIleHisGlnAsnGluAsnLy 2048

    4092 CTACCGCGCGGGCATTAAGGCAGATTATTCATTCAAACCGGCGCAACACA 4141
         :   :::||||||||||:::::                 |||:::   :
    2048 sAlaAsnAlaGlyIleSerSer.................AlaMetAlaM 2059

    4142 TTTCCATCACGCCTTATTTGAGCCTGTCCTATACCGATGCCGCTTCGGGC 4191
         :::::      |||         :::         :::   :::::|||
    2059 etAlaSerMetProGlnAlaTyrIleProGlyArgSerMetValThrGly 2075

    4192 AAAGTCCGAACACGCGTCAATACCGCCGTATTGGCTCAGGATTTCGGCAA 4241
            :::   |||:::   :::   :::::::::|||      ::::::|
    2076 GlyIleAlaThrHisAsnGlyGlnGlyAlaValAlaValGlyLeuSerLy 2092

    4242 A...ACCCGCAGTGCGGAATGG............GGCGTAAACGCCGAAA 4276
         |   :::   :::::::::|||         :::   :::|||:::
    2092 sLeuSerAspAsnGlyGlnTrpValPheLysIleAsnGlySerAlaAspT 2109

    4277 TCAAAGGTTTCACGCTGTCCCTCCACGCTGCCGCCGCCAAAGGC 4320
           :::|||              |||::::::||||||   |||
    2109 hrGlnGly..............HisValGlyAlaAlaValGly 2118

seq_name: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-409-995-4

seq_documentation_block:
; Sequence 4, Application US/08409995
; Patent No. 5646259
;  GENERAL INFORMATION:
;    APPLICANT:  Barenkamp, Stephen I.
;    APPLICANT:  St. Geme III, Joseph W.
;    TITLE OF INVENTION:  Haemophilus Adhesion Proteins
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Flehr, Hohbach, Test, Albritton & Herbert
;      STREET:  Four Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/409,995
;      FILING DATE:  24-MAR-1995
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Silva, Robin M.
;      REGISTRATION NUMBER:  38,304
;      REFERENCE/DOCKET NUMBER:  A-61053/RFT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1912 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
US-08-409-995-4


alignment_scores:
            Quality:  218.50              Length:   1492
              Ratio:  0.322                 Gaps:     73
 Percent Similarity:  45.442   Percent Identity: 19.035

alignment_block:
 US-09-303-518D-649 x US-08-409-995-4   ..

 Align seg 1/1  to: US-08-409-995-4  from: 1  to: 1912

     172 GCCGAAAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTA 221
         |||::::::   |||   :::         :::||||||   ||||||
     435 AlaAspThrAspGlyAlaLeuGluGlyIleSerLysAspGlnGluVal.. 450

     222 CAACAAAAAAGGGGAGTTGGTC............GGCAAATCAATG...A 256
             |||   ||||||   |||              ||||||::::::   :
     451 ....LysAlaGlyGluThrValThrPheLysAlaGlyLysAsnLeuLysV 466

     257 CAAAAGCCCCGATGATTGATTTTTCTGTGGTGTCGCGTAACGGCGTGGCG 306
         ::|||           :::|||:::         ::::::::::::::
     466 alLysGlnAspGlyAlaAsnPheThrTyrSerLeuGlnAspAlaLeuThr 482

     307 GCATTGGTGGGC..................................... 318
         :::|||::::::
     483 GlyLeuThrSerIleThrLeuGlyGlyThrThrAsnGlyGlyAsnAspAl 499

     319 ..................GATCAATATATTGTGAGCGTGGCACATAACG 349
                            |||         ::::::   |||   ||||
     499 aLysThrValIleAsnLysAspGlyLeuThrIleThrProAlaGlyAsnG 516

     350 GCGGCTATAACAACGTTGATTTTGGTGCGGAAGGAAGAAATCCCGATCAA 399
         ||                     |||:::   |||   |||
     516 ly.....................GlyThrThrGlyThrAsn......... 522

     400 CATCGTTTTACTTATAAAATTGTGAAACGGAATAATTATAAAGCAGGGAC 449
                 |||   :::::::::|||   ::::::   |||||||||::
     523 .........ThrIleSerValThrLys...AspGlyIleLysAlaGlyAs 535

     450 TAAAGGCCATCCTTATGGCGGCGATTATCATATGCCGCGTTTGCATAAAT 499
         :|||:::
     535 nLysAla........................................... 537

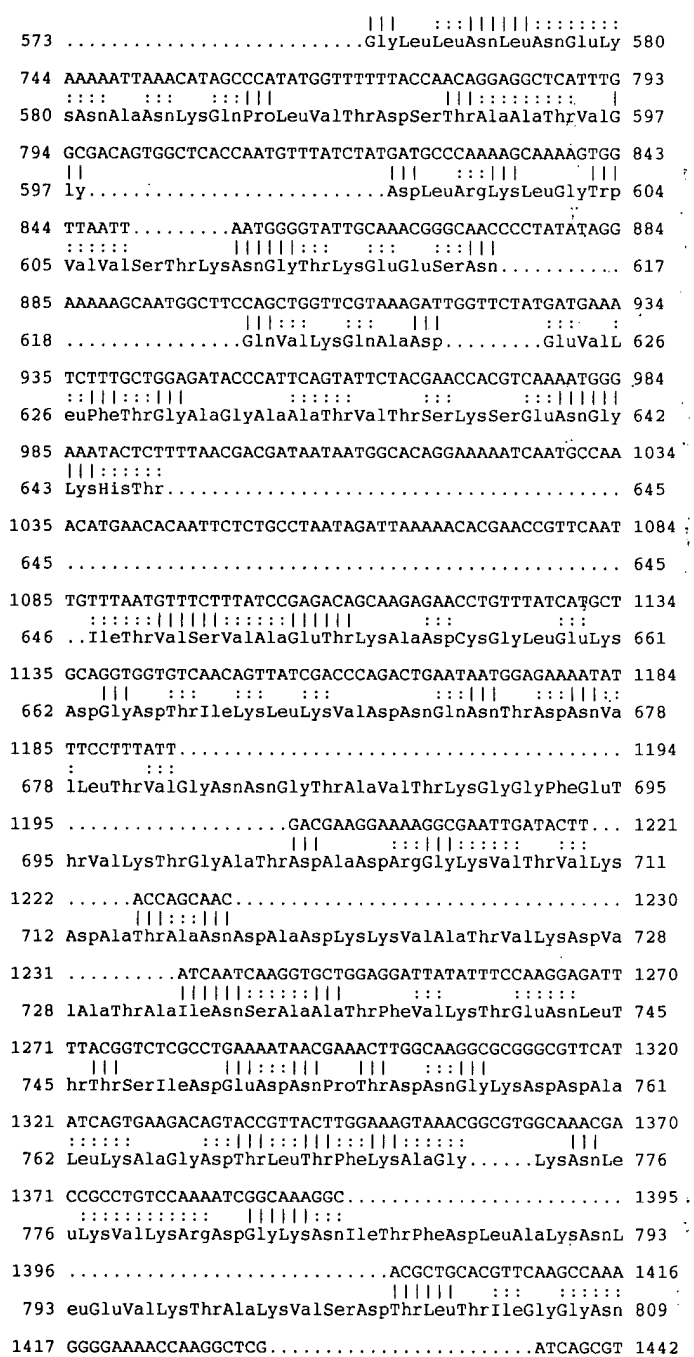
     500 TTGTCACAGATGCAGAACCTGTTGAAATGACCAGTTATATGGATGGGCGG 549
                                     :::|||:::      :::|||
     538 .........................IleThrAsnValAlaSerGlyLeu 545

     550 AAATATATCGATCAAAATAATTACCCTGACCGTGTTCGTATTGGGGCAGG 599
         :::      |||:::   |||:::
     546 ArgAlaTyrAspAspAlaAsnPhe........................... 553

     600 CAGGCAATATTGGCGATCTGATGAAGATGAGCCCAATAAC......CGCG 643
                                 |||      ||||||         :
     554 .........................AspValLeuAsnAsnSerAlaThrA 562

     644 AAAGTTCATATCATATTGCAAGTGCGTATTCTTGGCTCGTTGGTGGCAAT 693
         ::   :::   |||:::   :::||||||:::
     562 spLeuAsnArgHisValGluAspAlaTyrLys................... 572

     694 ACCTTTGCACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGTGA 743
```





```
2286 TGCCGATCACGCTCATTTAAATCTCACAGGGCTTGCCACACTCAACGGCA 2335
     |               :::|||      ||||||||||||      :::|
1088 u..............ValAsnAlaGluGlyLeuAlaThrAlaLeuAsnA 1100

2336 ATCTTAGT.......................GCAAATGGCGATACACGT 2361
     ||||||||                      |||:::|||:::::::::
1100 snLeuSerTrpThrAlaLysAlaAspLysTyrAlaAspGlyGluSerGlu 1116

2362 TATACAGTCAGCCACAACGCCACCCAAAACGGCAACCTTAGCCTC..... 2406
           ::::::::::::::::      :::   ::::::::::::
1117 GlyGluThrAspGlnGluValLysAlaGlyAspLysValThrPheLysAl 1133

2407 ...............GTGGGCAATGCCCAAGCAACATTTAAT....... 2433
                     |||   :::::::::      |||:::
1133 aGlyLysAsnLeuLysValLysGlnSerGluLysAspPheThrTyrSerL 1150

2434 ..CAAGCCACATTAAACGGCAACACATCGGCTTCG......GGCAATGCT 2475
       |||   ||||||:::|||   ||||||   :::      |||:::|||
1150 euGlnAspThrLeuThrGlyLeuThrSerIleThrLeuGlyGlyThrAla 1166

2476 TCATTTAATCTAAGCGACCACGCCGTACAAAACGGC.........AGTCT 2516
     :::   :::   :::       :::|||   |||            :::::
1167 AsnGlyArgAsnAspThrGlyThrValIleAsnLysAspGlyLeuThrIl 1183

2517 GACGCTTTCCGGCAACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACG 2566
     :||||||:::::::::|||   |||::::::      :::|||   ||||
1183 eThrLeuAlaAsnGlyAlaAlaAlaGlyThr......AspAlaSerAsnG 1198

2567 GTAAT.......................GTCTCCCTAGCCGATAAGGCA 2592
     |||||                      :::|||   ::::::|||
1198 lyAsnThrIleSerValThrLysAspGlyIleSerAlaGlyAsnLysGlu 1214

2593 GTATTCCATTTTGAAAGCAGCCGCTTTACCGGACAAATCAGCGGCGGCAA 2642
     :::   :::   :::|||:::      |||           :::      ::
1215 IleThrAsnValLysSerAlaLeuLysThr......TyrLysAspThrGl 1229

2643 GGATACGGCATTACACTTAAAAGACAGCGAATGG......ACGCTGCCGT 2686
     ::::||||||   :::   :::|||:::|||:::      ::::::   :
1229 nAsnThrAlaAspGluThrGlnAspLysGluPheHisAlaAlaValLysA 1246

2687 CAGGCACGGAATTAGGCAATTTAAACCTTGACAACGCCACCATTACA... 2733
     :::::::::|||:::      ::::::   ::::::||||||::::::
1246 snAlaAsnGluValGluPheValGlyLysAsnGlyAlaThrValSerAla 1262

2734 ...CTCAATTCCGCCTATCGCCAC............GATGCGGCAGGGGC 2768
        ::::::      :::|||            |||:::|||   ||
1263 LysThrAspAsnAsnGlyLysHisThrValThrIleAspValAlaGluAl 1279

2769 GCAAACCGGCAGTGCG............ACAGATGCG......CCGCGCC 2800
     |:::::::|||::::::           ||||||:::           :
1279 aLysValGlyAspGlyLeuGluLysAspThrAspGlyLysIleLysLeuL 1296

2801 GCCGTTCGCGCCGTTCGCGCCGTTCCCTATTATCCGTT......ACACCG 2844
     ::   ::::::   :::   ::::::|||||||:::|||      |||
1296 ysValAspAsnThrAspGlyAsnAsnLeuLeuThrValAspAlaThrLys 1312

2845 CCAACTTCGGTA...GAATCCCGTTTCAACACGCTGACG........... 2880
        :::||||||   :::::::::||||||::::::|||
1313 GlyAlaSerValAlaLysGlyGluPheAsnAlaValThrThrAspAlaTh 1329

2881 .............GTAAACGGCAAATTGAACGGTCAGGGAACATTCCGCT 2917
                  :::|||::::::   :::|||:::   :::   :::
1329 rThrAlaGlnGlyThrAsnAlaAsnGluArgGlyLysValValValLysG 1346

2918 TTATGTCGGAACTCTTCGGCTACCGCAGCGACAAATTGAAGCTGGCGGAA 2967
        :::          :::   ::::::||||||   |||:::|||
1346 lySerAsnGlyAlaThrAlaThrGluThrAspLysLysLysValAla... 1361

2968 AGTTCCGAAGGCACTTACACCTTGGCGGTCAACAATACCGGCAAC..... 3012
        :::   |||      :::   |||:::|||:::::::::::::
1362 ...ThrValGlyAspValAlaLysAlaIleAsnAspAlaAlaThrPheVa 1377

3013 .GAACCTGCAAGCCTCGAACAATTGACGGTAGTGGAAGGAAAAGACAACA 3061
      :::      :::   ::::::   |||:::           |||::::
1377 lLysValGluAsnAspAspSerAlaThrIle...........AspAspS 1390

3062 AACCGCTGTCCGAAAACCTTAATTTCACCCTGCAA...AACGAACACGTC 3108
     ::|||   :::::::::   |||   :::|||:::      :::   :::
1390 erProThrAspAspGlyAlaAsnAspAlaLeuLysAla***AspThrLeu 1406

3109 GATGCCGGCGCGTGGCGTTACCAACTCATCCGCAAAGACGGCGAG..... 3153
               |||   :::         :::::::::||||||:::
1407 ThrLeuLysAlaGlyLysAsnLeuLysValLysArgAspGlyLysAsnIl 1423

3154 ....TTCCGCCTGCATAAT......CCGGTCAAAGAACAAGAGCTTTCCG 3193
         |||   |||   |||         ||||||:::      :::||||
1423 eThrPheAlaLeuAlaAsnAspLeuSerValLysSerAlaThrValSerA 1440

3194 ACAAACTC.................GGCAAGGCAGAAGCCAAAAAACAG 3225
     ||||||||                ::::|||::::::      ::::::
1440 spLysLeuSerLeuGlyThrAsnGlyAsnLysValAsnIleThrSerAsp 1456

3226 GCGGAA..............AAAGACAACGCGCAAAGCCTTGACGCGCT 3260
     ::::::               ||||||:::      :::   ||||||
1457 ThrLysGlyLeuAsnPheAlaLysAspSerLysThrGlyAspAspAlaAs 1473

3261 GATTGCGGCCGGGCGCGATGCCGTCGAAAAGACAGAAAGCGTT......G 3304
      |||      :::      |||      |||:::::::::      :
1473 nIleHisLeuAsnGlyIleAlaSerThrLeuThrAspThrLeuLeuAsnS 1490

3305 CCGAACCGGCCCGGCAGGCAGGCGGGGAAAATGTCGGCATTATGCAGGCG 3354
     ::      :::   :::   ||||||:::::::::            :::
1490 erGlyAlaThrThrAsnLeuGlyGlyAsnGlyIle.....Thr 1502

3355 GAGGAAGAGAAAAAACGGGTGCAGGCGGATAAAGACACCGCCTTGGCGAA 3404
     ::::::|||||||||   :::   :::   ||||||:::      |||
1503 AspAsnGluLysLys***AlaAlaSerValLysAspValLeuAsnAlaGl 1519

3405 ACAGCGCGAAGCGGAAACCCGGCCGGCTACCACCGCCTTCCCCCGCGCCC 3454
        :::          ::::::||||||:::
1519 yTrpAsnValArgGlyValLysProAlaSer..................... 1529

3455 GCCGCGCCCGCCGGGATTTGCCGCAACTGCAACCCCAACCGCAGCCCCAA 3504

1529 ..................................................... 1529

3505 CCGCAGCGCGACCTGATCAGCCGTTATGCCAATAGCGGTTTGAGT..... 3549
                                 |||||||:::   ::::::
1530 ..........................AlaAsnAsnGlnValGluAsnIl 1537

3550 .GAATTTTCCGCCACGCTCAACAGCGTTTTCGCCGTACAGGACGAATTAG 3598
      :::|||   ||||||   :::::::|||      |||:::   :::   |
1537 eAspPheValAlaThrTyrAspThrValAspPheValSerGlyAspLysA 1554

3599 AC......CGCGTATTTGCCGAAGACCGCCGCAAC.............. 3627
     ||          |||   :::|||::::::   |||
1554 spThrThrSerValThrValGluSerLysAspAsnGlyLysArgThrGlu 1570

3628 .........GCCGTTTGGACAAGCGGCATCCGGGACACCAAACACTACCG 3668
             ::::::   ||||||   |||:::|||         :::::
1571 ValLysIleGlyAlaLysThrSerValIleLysAsp.........HisAs 1584

3669 TTCGCAAGATTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCAAATCG 3718
     :::::::   |||   :::   ::::::      |||   ::::::   |
1584 nGlyLysLeuPheThrGly...LysGluLeuLysAspAlaAsnAsnAsnG 1600

3719 GTATGCAG..................AAAAACCTCGGCAGCGGGCGCGTC 3750
     ||:::                    |||:::   |||:::|||   |||
1600 lyValThrValThrGluThrAspGlyLysAspGluGlyAsnGlyLeuVal 1616

3751 .........GGCATCCTGTTTTCGCACAACCGG.................. 3774
             :::::::::   :::   |||:::
1617 ThrAlaLysAlaValIleAspAlaValAsnLysAlaGlyTrpArgValLy 1633

3774 ..................................................... 3774
```

```
    1633 sThrThrGlyAlaAsnGlyGlnAsnAspAspPheAlaThrValAlaSerG 1650

    3775 ..ACCGAAAACACCTTCGACGACGGCATCGGCAACTCGGCACGGCTTGCC 3822
           |||:::   ||||||   ||||||   |||:::::|||:::::::::
    1650 lyThrAsnValThrPheAlaAspGlyAsnGlyThrThrAlaGluValThr 1666

    3823 .........CACGGCGCCGTTTTCGGGCAATACGGCATCGACAGGTTCTA 3863
                   |||::::::      :::|||::::::
    1667 LysAlaAsnAspGlySerIleThrValLysTyrAsnVal........... 1679

    3864 CATCGGCATCAGCGCGGGCGCGGGTTTTAGCAGCGGCAGCCTTTCAGACG 3913
                                              :::::::::||||
    1680 .....................................LysValAlaAspG 1684

    3914 GCATC.........GGAGGCAAAATC 3930
         ||:::         |||   ||||||
    1684 lyLeuLysLeuAspGlyAspLysIle 1692

seq_name: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:US-08-685-467-4

seq_documentation_block:
; Sequence 4, Application US/08685467
; Patent No. 6060059
;  GENERAL INFORMATION:
;    APPLICANT:  St. Geme III, Joseph W.
;    APPLICANT:  Barenkamp, Stephen J.
;    TITLE OF INVENTION:  HAEMOPHILUS ADHESION PROTEINS
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Flehr, Hohbach, Test, Albritton & Herbert
;      STREET:  Four Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/685,467
;      FILING DATE:  22-JUL-1996
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/409,995
;      FILING DATE:  24-MAR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Silva, Robin M.
;      REGISTRATION NUMBER:  38,304
;      REFERENCE/DOCKET NUMBER:  A-61053-2/RFT/RMS
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1912 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-08-685-467-4


alignment_scores:
            Quality:  218.50              Length:   1492
              Ratio:  0.322                 Gaps:     73
 Percent Similarity:  45.442   Percent Identity: 19.035

alignment_block:
 US-09-303-518D-649 x US-08-685-467-4   ..

Align seg 1/1  to: US-08-685-467-4  from: 1  to: 1912

    172 GCCGAAAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTA 221
        |||::::::   |||   :::          :::||||||   ||||||
    435 AlaAspThrAspGlyAlaLeuGluGlyIleSerLysAspGlnGluVal.. 450

    222 CAACAAAAAAGGGGAGTTGGTC............GGCAAATCAATG...A 256
            |||   ||||||   |||            ||||||::::::   :
    451 ....LysAlaGlyGluThrValThrPheLysAlaGlyLysAsnLeuLysV 466

    257 CAAAAGCCCCGATGATTGATTTTTCTGTGGTGTCGCGTAACGGCGTGGCG 306
        ::|||          :::|||:::          ::::::::::::::
    466 alLysGlnAspGlyAlaAsnPheThrTyrSerLeuGlnAspAlaLeuThr 482

    307 GCATTGGTGGGC..................................... 318
        :::|||::::::
    483 GlyLeuThrSerIleThrLeuGlyGlyThrThrAsnGlyGlyAsnAspAl 499

    319 ..................GATCAATATATTGTGAGCGTGGCACATAACG 349
                           |||         ::::::   |||   ||||
    499 aLysThrValIleAsnLysAspGlyLeuThrIleThrProAlaGlyAsnG 516

    350 GCGGCTATAACAACGTTGATTTTGGTGCGGAAGGAAGAAATCCCGATCAA 399
        ||                     |||:::   |||   |||
    516 ly.....................GlyThrThrGlyThrAsn......... 522

    400 CATCGTTTTACTTATAAAATTGTGAAACGGAATAATTATAAAGCAGGGAC 449
                |||   ::::::::::|||   ::::::   ||||||||::
    523 .........ThrIleSerValThrLys...AspGlyIleLysAlaGlyAs 535

    450 TAAAGGCCATCCTTATGGCGGCGATTATCATATGCCGCGTTTGCATAAAT 499
        :|||::.
    535 nLysAla.......................................... 537

    500 TTGTCACAGATGCAGAACCTGTTGAAATGACCAGTTATATGGATGGGCGG 549
                                     :::|||:::      :::|||
    538 ..........................IleThrAsnValAlaSerGlyLeu 545

    550 AAATATATCGATCAAAATAATTACCCTGACCGTGTTCGTATTGGGGCAGG 599
        :::      |||:::   |||:::
    546 ArgAlaTyrAspAspAlaAsnPhe........................... 553

    600 CAGGCAATATTGGCGATCTGATGAAGATGAGCCCAATAAC......CGCG 643
                                  |||      ||||||         :
    554 .........................AspValLeuAsnAsnSerAlaThrA 562

    644 AAAGTTCATATCATATTGCAAGTGCGTATTCTTGGCTCGTTGGTGGCAAT 693
        ::   :::   |||:::   :::|||||||:::
    562 spLeuAsnArgHisValGluAspAlaTyrLys.................. 572

    694 ACCTTTGCACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGTGA 743
                                     |||   :::|||||||:::::::
    573 ..........................GlyLeuLeuAsnLeuAsnGluLy 580

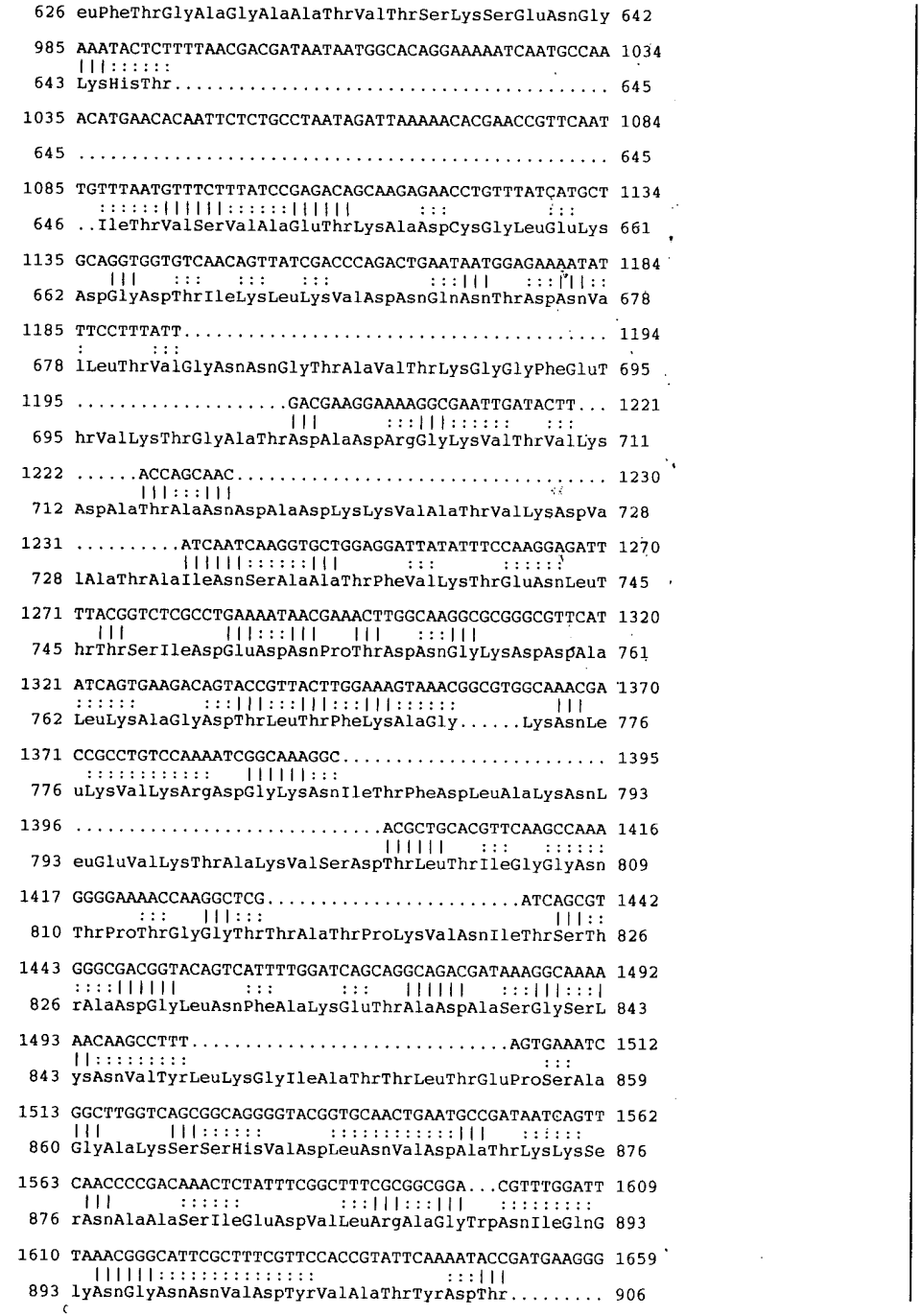
    744 AAAAATTAAACATAGCCCATATGGTTTTTTACCAACAGGAGGCTCATTTG 793
        ::::   :::   :::|||               |||:::::::::   |
    580 sAsnAlaAsnLysGlnProLeuValThrAspSerThrAlaAlaThrValG 597

    794 GCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGG 843
        ||                        |||   :::|||      |||
    597 ly.........................AspLeuArgLysLeuGlyTrp 604

    844 TTAATT.........AATGGGGTATTGCAAACGGGCAACCCCTATATAGG 884
        ::::::         ||||||:::   :::   :::|||
    605 ValValSerThrLysAsnGlyThrLysGluGluSerAsn........... 617

    885 AAAAAGCAATGGCTTCCAGCTGGTTCGTAAAGATTGGTTCTATGATGAAA 934
                      |||:::   :::   |||         :::   :
    618 ...............GlnValLysGlnAlaAsp.........GluValL 626

    935 TCTTTGCTGGAGATACCCATTCAGTATTCTACGAACCACGTCAAAATGGG 984
        ::|||:::|||         ::::::      :::      :::||||||
```


626 euPheThrGlyAlaGlyAlaAlaThrValThrSerLysSerGluAsnGly 642
985 AAATACTCTTTTAACGACGATAATAATGGCACAGGAAAAATCAATGCCAA 1034
643 LysHisThr..... 645
1035 ACATGAACACAATTCTCTGCCTAATAGATTAAAAACACGAACCGTTCAAT 1084
645 645
1085 TGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTGTTTATCATGCT 1134
646 ..IleThrValSerValAlaGluThrLysAlaAspCysGlyLeuGluLys 661
1135 GCAGGTGGTGTCAACAGTTATCGACCCAGACTGAATAATGGAGAAAATAT 1184
662 AspGlyAspThrIleLysLeuLysValAspAsnGlnAsnThrAspAsnVa 678
1185 TTCCTTTATT..... 1194
678 lLeuThrValGlyAsnAsnGlyThrAlaValThrLysGlyGlyPheGluT 695
1195 GACGAAGGAAAAGGCGAATTGATACTT... 1221
695 hrValLysThrGlyAlaThrAspAlaAspArgGlyLysValThrValLys 711
1222 ACCAGCAAC..... 1230
712 AspAlaThrAlaAsnAspAlaAspLysLysValAlaThrValLysAspVa 728
1231 ATCAATCAAGGTGCTGGAGGATTATATTTCCAAGGAGATT 1270
728 lAlaThrAlaIleAsnSerAlaAlaThrPheValLysThrGluAsnLeuT 745
1271 TTACGGTCTCGCCTGAAAATAACGAAACTTGGCAAGGCGCGGGCGTTCAT 1320
745 hrThrSerIleAspGluAspAsnProThrAspAsnGlyLysAspAspAla 761
1321 ATCAGTGAAGACAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGA 1370
762 LeuLysAlaGlyAspThrLeuThrPheLysAlaGly......LysAsnLe 776
1371 CCGCCTGTCCAAAATCGGCAAAGGC..... 1395
776 uLysValLysArgAspGlyLysAsnIleThrPheAspLeuAlaLysAsnL 793
1396 ACGCTGCACGTTCAAGCCAAA 1416
793 euGluValLysThrAlaLysValSerAspThrLeuThrIleGlyGlyAsn 809
1417 GGGGAAAACCAAGGCTCG.....ATCAGCGT 1442
810 ThrProThrGlyGlyThrThrAlaThrProLysValAsnIleThrSerTh 826
1443 GGGCGACGGTACAGTCATTTTGGATCAGCAGGCAGACGATAAAGGCAAAA 1492
826 rAlaAspGlyLeuAsnPheAlaLysGluThrAlaAspAlaSerGlySerL 843
1493 AACAAGCCTTT.....AGTGAAATC 1512
843 ysAsnValTyrLeuLysGlyIleAlaThrThrLeuThrGluProSerAla 859
1513 GGCTTGGTCAGCGGCAGGGGTACGGTGCAACTGAATGCCGATAATCAGTT 1562
860 GlyAlaLysSerSerHisValAspLeuAsnValAspAlaThrLysLysSe 876
1563 CAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGA...CGTTTGGATT 1609
876 rAsnAlaAlaSerIleGluAspValLeuArgAlaGlyTrpAsnIleGlnG 893
1610 TAAACGGGCATTCGCTTTCGTTCCACCGTATTCAAAATACCGATGAAGGG 1659
893 lyAsnGlyAsnAsnValAspTyrValAlaThrTyrAspThr..... 906
1660 GCGATGATTGTCAACCACAATCAAGACAAAGAATCCACCGTTACCATTAC 1709
907 ValAsnPheThrAspAspSerThrGlyThrThrThrValTh 920
1710 AGGCAATAAAGATATTGCTACAACCGGCAATAACAACAGCTTGGATAGCA 1759
920 rValThrGlnLysAlaAspGlyLysGlyAlaAspValLysIleGlyAlaL 937
1760 AAAAAGAAATTGCC......TACAACGGT...TGGTTTGGCGAGAAAGAT 1800
937 ysThrSerValIleLysAspHisAsnGlyLysLeuPheThrGlyLysAsp 953
1801 ACGACCAAAACGAACGGGCGGCTCAACCTTGTTTACCAGCCCGCCGCAGA 1850
954 LeuLysAspAlaAsnAsnGlyAlaThrValSerGluAspAspGlyLysAs 970
1851 AGACCGCACCCTGCTGCTTTCCGGCGGAACA.....AATT 1885
970 pThrGlyThrGlyLeuValThrAlaLysThrValIleAspAlaValAsnL 987
1886 TAAACGGCAACATCACGCAAACAAACGGCAAACTGTTTTTCAGCGGCAGA 1935
987 ysSerGlyTrpArgValThrGlyGluGlyAlaThrAlaGluThrGlyAla 1003
1936 CCAACACCGCACGCCTACAATCATTTAAACGACCATTGGTCGCAAAAAGA 1985
1004 ThrAlaValAsnAlaGlyAsn.....AlaGl 1012
1986 GGGCATTCCTCGCGGGGAAATCGTGTGGGACAACGACTGGATCAACCGCA 2035
1012 uThrValThrSerGlyThrSerVal..... 1020
2036 CATTTAAAGCGGAAAACTTCCAAATTAAAGGCGGACAGGCGGTGGTTTCC 2085
1021 AsnPhe.....LysAsnGlyAsnAlaThrThrAla 1030
2086 CGCAATGTTGCCAAAGTGAAAGGCGATTGGCATTTGAGCAATCACGCCCA 2135
1031 ...ThrValSerLysAspAsnGlyAsnIleAsnValLysTyrAspValAs 1046
2136 AGCAGTTTTTGGTGTCGCACCGCATCAAAGCCACACAATCTGTACACGTT 2185
1046 nValGlyAspGlyLeuLysIleGlyAspAspLysLysIleVal.....A 1061
2186 CGGACTGGACGGGTCTGACAAATTGTGTCGAAAAAACCATTACCGACGAT 2235
1061 laAspThrThrThrLeu.....ThrValThrGlyGly 1071
2236 AAAGTGATTGCTTCATTGACTAAGACCGACATCAGCGGCAATGTCGATCT 2285
1072 LysValSerValProAlaGlyAlaAsnSerValAsnAsnAsnLysLysLe 1088
2286 TGCCGATCACGCTCATTTAAATCTCACAGGGCTTGCCACACTCAACGGCA 2335
1088 u.....ValAsnAlaGluGlyLeuAlaThrAlaLeuAsnA 1100
2336 ATCTTAGT.....GCAAATGGCGATACACGT 2361
1100 snLeuSerTrpThrAlaLysAlaAspLysTyrAlaAspGlyGluSerGlu 1116
2362 TATACAGTCAGCCACAACGCCACCCAAAACGGCAACCTTAGCCTC..... 2406
1117 GlyGluThrAspGlnGluValLysAlaGlyAspLysValThrPheLysAl 1133
2407 GTGGGCAATGCCCAAGCAACATTTAAT..... 2433
1133 aGlyLysAsnLeuLysValLysGlnSerGluLysAspPheThrTyrSerL 1150
2434 ..CAAGCCACATTAAACGGCAACACATCGGCTTCG.....GGCAATGCT 2475
1150 euGlnAspThrLeuThrGlyLeuThrSerIleThrLeuGlyGlyThrAla 1166

```
2476 TCATTTAATCTAAGCGACCACGCCGTACAAAACGGC.........AGTCT 2516
     :::   :::   :::       :::|||   |||             :::::
1167 AsnGlyArgAsnAspThrGlyThrValIleAsnLysAspGlyLeuThrIl 1183

2517 GACGCTTTCCGGCAACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACG 2566
     :||||||:::::::::|||   |||:::::      :::|||   ||||
1183 eThrLeuAlaAsnGlyAlaAlaAlaGlyThr......AspAlaSerAsnG 1198

2567 GTAAT.......................GTCTCCCTAGCCGATAAGGCA 2592
     |||||                      :::|||   :::::::|||
1198 lyAsnThrIleSerValThrLysAspGlyIleSerAlaGlyAsnLysGlu 1214

2593 GTATTCCATTTTGAAAGCAGCCGCTTTACCGGACAAATCAGCGGCGGCAA 2642
     :::   :::   :::|||:::      |||         :::      ::
1215 IleThrAsnValLysSerAlaLeuLysThr......TyrLysAspThrGl 1229

2643 GGATACGGCATTACACTTAAAAGACAGCGAATGG......ACGCTGCCGT 2686
     ::::||||||   :::   :::|||:::|||:::      ::::::   :
1229 nAsnThrAlaAspGluThrGlnAspLysGluPheHisAlaAlaValLysA 1246

2687 CAGGCACGGAATTAGGCAATTTAAACCTTGACAACGCCACCATTACA... 2733
     ::::::::|||:::      ::::::   ::::::||||||::::::
1246 snAlaAsnGluValGluPheValGlyLysAsnGlyAlaThrValSerAla 1262

2734 ...CTCAATTCCGCCTATCGCCAC............GATGCGGCAGGGGC 2768
          ::::::      :::|||            |||:::|||   ||
1263 LysThrAspAsnAsnGlyLysHisThrValThrIleAspValAlaGluAl 1279

2769 GCAAACCGGCAGTGCG............ACAGATGCG......CCGCGCC 2800
     |:::::|||:::::             ||||||:::           :
1279 aLysValGlyAspGlyLeuGluLysAspThrAspGlyLysIleLysLeuL 1296

2801 GCCGTTCGCGCCGTTCGCGCCGTTCCCTATTATCCGTT......ACACCG 2844
     ::   ::::::   :::   ::::::||||||:::|||      |||
1296 ysValAspAsnThrAspGlyAsnAsnLeuLeuThrValAspAlaThrLys 1312

2845 CCAACTTCGGTA...GAATCCCGTTTCAACACGCTGACG........... 2880
        :::||||||   :::::::::||||||:::::|||
1313 GlyAlaSerValAlaLysGlyGluPheAsnAlaValThrThrAspAlaTh 1329

2881 .............GTAAACGGCAAATTGAACGGTCAGGGAACATTCCGCT 2917
                  :::|||::::::   :::|||:::   :::   :::
1329 rThrAlaGlnGlyThrAsnAlaAsnGluArgGlyLysValValValLysG 1346

2918 TTATGTCGGAACTCTTCGGCTACCGCAGCGACAAATTGAAGCTGGCGGAA 2967
        :::         :::   ::::::||||||   |||:::|||
1346 lySerAsnGlyAlaThrAlaThrGluThrAspLysLysLysValAla... 1361

2968 AGTTCCGAAGGCACTTACACCTTGGCGGTCAACAATACCGGCAAC..... 3012
     :::   |||      :::   |||:::|||:::::::::::
1362 ...ThrValGlyAspValAlaLysAlaIleAsnAspAlaAlaThrPheVa 1377

3013 .GAACCTGCAAGCCTCGAACAATTGACGGTAGTGGAAGGAAAAGACAACA 3061
      :::      :::   ::::::   |||:::            |||::::
1377 lLysValGluAsnAspAspSerAlaThrIle............AspAspS 1390

3062 AACCGCTGTCCGAAAACCTTAATTTCACCCTGCAA...AACGAACACGTC 3108
     ::|||   :::::::::   |||   :::|||:::      :::   :::
1390 erProThrAspAspGlyAlaAsnAspAlaLeuLysAla***AspThrLeu 1406

3109 GATGCCGGCGCGTGGCGTTACCAACTCATCCGCAAAGACGGCGAG..... 3153
               |||   :::         :::::::::||||||:::
1407 ThrLeuLysAlaGlyLysAsnLeuLysValLysArgAspGlyLysAsnIl 1423

3154 ....TTCCGCCTGCATAAT......CCGGTCAAAGAACAAGAGCTTTCCG 3193
        |||   |||   |||         ||||||:::      :::||||
1423 eThrPheAlaLeuAlaAsnAspLeuSerValLysSerAlaThrValSerA 1440

3194 ACAAACTC.................GGCAAGGCAGAAGCCAAAAAACAG 3225
     ||||||||                :::|||:::::      ::::::
1440 spLysLeuSerLeuGlyThrAsnGlyAsnLysValAsnIleThrSerAsp 1456

3226 GCGGAA..............AAAGACAACGCGCAAAGCCTTGACGCGCT 3260
```

```
          ::::::               ||||||:::      :::   ||||||
1457 ThrLysGlyLeuAsnPheAlaLysAspSerLysThrGlyAspAspAlaAs 1473

3261 GATTGCGGCCGGGCGCGATGCCGTCGAAAAGACAGAAAGCGTT......G 3304
      |||      :::      |||         |||:::::::::      :
1473 nIleHisLeuAsnGlyIleAlaSerThrLeuThrAspThrLeuLeuAsnS 1490

3305 CCGAACCGGCCCGGCAGGCAGGCGGGGAAAATGTCGGCATTATGCAGGCG 3354
     ::      :::   :::   ||||||::::::::::             :::
1490 erGlyAlaThrThrAsnLeuGlyGlyAsnGlyIle............Thr 1502

3355 GAGGAAGAGAAAAAACGGGTGCAGGCGGATAAAGACACCGCCTTGGCGAA 3404
     ::::::||||||||||   :::   :::   |||||||:::      |||
1503 AspAsnGluLysLys***AlaAlaSerValLysAspValLeuAsnAlaGl 1519

3405 ACAGCGCGAAGCGGAAACCCGGCCGGCTACCACCGCCTTCCCCCGCGCCC 3454
         :::         ::::::||||||:::
1519 yTrpAsnValArgGlyValLysProAlaSer................... 1529

3455 GCCGCGCCCGCCGGGATTTGCCGCAACTGCAACCCCAACCGCAGCCCCAA 3504

1529 ..................................................... 1529

3505 CCGCAGCGCGACCTGATCAGCCGTTATGCCAATAGCGGTTTGAGT..... 3549
                                  ||||||:::   ::::::
1530 ........................AlaAsnAsnGlnValGluAsnIl 1537

3550 .GAATTTTCCGCCACGCTCAACAGCGTTTTCGCCGTACAGGACGAATTAG 3598
      :::|||   ||||||   :::::|||      |||:::   :::   |
1537 eAspPheValAlaThrTyrAspThrValAspPheValSerGlyAspLysA 1554

3599 AC......CGCGTATTTGCCGAAGACCGCCGCAAC............... 3627
     ||         |||   :::|||::::::   |||
1554 spThrThrSerValThrValGluSerLysAspAsnGlyLysArgThrGlu 1570

3628 .........GCCGTTTGGACAAGCGGCATCCGGGACACCAAACACTACCG 3668
             :::::::   ||||||   |||:::|||         :::::
1571 ValLysIleGlyAlaLysThrSerValIleLysAsp.........HisAs 1584

3669 TTCGCAAGATTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCAAATCG 3718
     :::::::   |||   :::   ::::::      |||   :::::   |
1584 nGlyLysLeuPheThrGly...LysGluLeuLysAspAlaAsnAsnAsnG 1600

3719 GTATGCAG.................AAAAACCTCGGCAGCGGGCGCGTC 3750
     ||:::                    |||:::   |||:::|||   |||
1600 lyValThrValThrGluThrAspGlyLysAspGluGlyAsnGlyLeuVal 1616

3751 .........GGCATCCTGTTTTCGCACAACCGG.................. 3774
             :::::::::   :::   |||:::
1617 ThrAlaLysAlaValIleAspAlaValAsnLysAlaGlyTrpArgValLy 1633

3774 ..................................................... 3774

1633 sThrThrGlyAlaAsnGlyGlnAsnAspAspPheAlaThrValAlaSerG 1650

3775 ..ACCGAAAACACCTTCGACGACGGCATCGGCAACTCGGCACGGCTTGCC 3822
       |||:::   ||||||   ||||||   |||::::::|||::::::::::
1650 lyThrAsnValThrPheAlaAspGlyAsnGlyThrThrAlaGluValThr 1666

3823 .........CACGGCGCCGTTTTCGGGCAATACGGCATCGACAGGTTCTA 3863
                |||::::::      :::|||::::::
1667 LysAlaAsnAspGlySerIleThrValLysTyrAsnVal............ 1679

3864 CATCGGCATCAGCGCGGGCGCGGGTTTTAGCAGCGGCAGCCTTTCAGACG 3913
                                          :::::::::||||
1680 ................................LysValAlaAspG 1684

3914 GCATC.........GGAGGCAAAATC 3930
     ||:::         |||   ||||||
1684 lyLeuLysLeuAspGlyAspLysIle 1692
```

seq_name: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-409-995-2

```
seq_documentation_block:
; Sequence 2, Application US/08409995
; Patent No. 5646259
;  GENERAL INFORMATION:
;    APPLICANT:  Barenkamp, Stephen I.
;    APPLICANT:  St. Geme III, Joseph W.
;    TITLE OF INVENTION:  Haemophilus Adhesion Proteins
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Flehr, Hohbach, Test, Albritton & Herbert
;      STREET:  Four Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/409,995
;      FILING DATE:  24-MAR-1995
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Silva, Robin M.
;      REGISTRATION NUMBER:  38,304
;      REFERENCE/DOCKET NUMBER:  A-61053/RFT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1098 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
US-08-409-995-2


alignment_scores:
            Quality:  214.00            Length:   1396
              Ratio:  0.347               Gaps:     72
 Percent Similarity:  44.198   Percent Identity: 20.129

alignment_block:
 US-09-303-518D-649 x US-08-409-995-2   ..

 Align seg 1/1  to: US-08-409-995-2  from: 1  to: 1098

      22 ACAACCGAAACACACCGCAAAGCCCCGAAAACCGGCCGCATCCGCTTCTC 71
         |||:::|||:::::::::::::::::|||   |||:::::::::::::
      46 ThrValGluAlaAsnAsnAsnThrProValThrAsnLysLeuLys..... 60

      72 GCCTGCTTACTTAGCCATATGCCTGTCGTTCGGCATTCTTCCCCAAGCCT 121
            ||||||
      61 ....AlaTyr..................................... 62

     122 GGGCGGGACACACTTATTTCGGCATCAACTACCAATACTATCGCGACTTT 171
            |||   :::   |||:::::::::   :::         |||
      63 .....GlyAspAlaAsnPheAsnPheThrAsnAsnSerIleAlaAspAla 77

     172 GCCGAAAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTA 221
          ::::::     :::         |||      :::::::::
      78 GluLysGlnValGlnGluAlaTyrLysGlyLeuLeuAsnLeuAsnGluLy 94

     222 CAACAAAAAAGGGGAGTTGGTCGGCAAATCAATGACAAAAGCCCCGATG. 270
          |||   :::   :::|||:::   ::::::   |||   |||   :::
      94 sAsnAlaSerAspLysLeuLeuValGluAspAsnThrAlaAlaThrValG 111

     271 .............ATTGATTTTTCTGTGGTGTCGCGTAACGGCGTGGCG 306
                      :::   :::   :::   |||:::||||||:::

     111 lyAsnLeuArgLysLeuGlyTrpValLeuSerSerLysAsnGlyThrArg 127

     307 GCATTGGTGGGCGATCAATATATTGTGAGCGTGGCACAT.......... 345
              ::::::|||           |||   |||
     128 AsnGluLysSerGlnGln.............ValLysHisAlaAspGluVa 140

     346 .............AACGGCGGCTATAACAACGTTGATTTTGGTGCGGAAG 382
                      :::||||||   :::   ::::::   :::   :::|
     140 lLeuPheGluGlyLysGlyGlyValGlnValThrSerThrSerGluAsnG 157

     383 GAAGAAATCCCGATCAACATCGTTTTACTTATAAAATTGTGAAACGGAAT 432
         ||:::            |||   :::|||:::   ::::::|||
     157 lyLys............HisThrIleThrPheAlaLeuAlaLysAspLeu 169

     433 AATTATAAAGCAGGGACT.....................AAAGGCCATCC 461
         :::   |||::::::|||                        |||
     170 GlyValLysThrAlaThrValSerAspThrLeuThrIleGlyGlyGlyAl 186

     462 TTATGGCGGCGATTATCATATGCCGCGTTTGCATAAATTTGTCACAGATG 511
             :::|||                          :::|||   :
     186 aAlaAlaGly............................AlaThrThrT 193

     512 CAGAACCTGTTGAAATGACCAGTTATATGGATGGGCGGAAATATATCGAT 561
         ::      |||::::::|||||      ||||||   |||:::
     193 hrProLysValAsnValThrSerThrThrAspGlyLeuLysPhe...... 207

     562 CAAAATAATTACCCTGACCGTGTTCGTATTGGGGCAGGCAGGCAATATTG 611
                              ::::::   :::||||||
     208 .....................AlaLysAspAlaAlaGlyAla........ 214

     612 GCGATCTGATGAAGATGAGCCCAATAACCGCGAAAGTTCATATCATATTG 661
                                  |||   :::::::::   |||:::
     215 ........................AsnGlyAspThrThrValHisLeu. 222

     662 CAAGTGCGTATTCTTGGCTCGTTGGTGGCAATACCTTTGCACAAAATGGA 711
                                                      ||||||
     223 ..............................................AsnGly 224

     712 TCAGGTGGTGGCACAGTCAACTTAGGTAGTGAAAAAATTAAACATAGCCC 761
            |||:::                  ::::::   :::      |||||
     225 IleGlySerThrLeu............ThrAspThrLeuValGlySerPr 237

     762 ATATGGTTTTTTACCAACAGGAGGCTCATTTGGCGACAGTGGCTCACCAA 811
         |      :::      |||   :::
     237 oAlaThrHisIleAspGlyGlyAspGlnSerThrHisTyrThrArgAlaA 254

     812 TGTTTATCTATGATGCCCAAAAGCAAAAGTGGTTAATTAATGGGGTATTG 861
              |||   |||:::   :::      |||   |||:::||||||
     254 laSerIleLysAspValLeuAsnAlaGlyTrpAsnIleLysGlyVal... 269

     862 CAAACGGGCAACCCCTATATAGGAAAAAGCAATGGCTTCCAGCTGGTTCG 911
         ::::::|||:::         |||:::|||::::::   ::::::|||::
     270 LysAlaGlySerThr...ThrGlyGlnSerGluAsnValAspPheValHi 285

     912 TAAAGATTGGTTCTATGAT......GAAATCTTTGCTGGAGATACCCAT. 954
         :              ||||||      |||:::::::::::::||||||:::
     285 sThr.........TyrAspThrValGluPheLeuSerAlaAspThrGluT 299

     955 .....TCAGTATTCTACGAACCACGTCAAAATGGGAAATACTCTTTTAAC 999
              :::|||      :::   ::::::|||||||||   :::
     299 hrThrThrValThrValAspSerLysGluAsnGlyLysArgThr...... 313

    1000 GACGATAATAATGGCACAGGAAAAATCAATGCCAAACATGAACACAATTC 1049
                                |||||||:::||||||
     314 ..............GluValLysIleGlyAlaLys.............. 320

    1050 TCTGCCTAATAGATTAAAAACACGAACCGTTCAATTGTTTAATGTTTCTT 1099
                  :::   :::|||   :::      :::||||||
     321 ....ThrSerValIleLysGluLysAspGlyLysLeuPhe.......... 332

    1100 TATCCGAGACAGCAAGAGAACCTGTTTATCATGCTGCAGGTGGTGTCAAC 1149
                                            :::|||   :::|||
     333 ................................ThrGlyLysAlaAsn 337
```



```
      ::::::          ||||||          :::|||   :::   |||
 789 GlnVal......AlaAspAlaIleAlaLysSerGlyPheGluLeuGlyLe 803

2781 TGCGACAGATGCGCCGCGCCGCCGTTCGCGCCGTTCGCGCCGTTCCCTAT 2830
      |||   :::|||      :::|||:::      ::::::   :::   |
 803 uAlaAspGluAlaAspAlaLysArgAlaPheAspAspLysThrLysAlaL 820

2831 TATCCGTTACACCGCCAACTTCGGTAGAATCCCGTTTCAACACGCTGACG 2880
     |||||:::
 820 euSerAla..................................... 822

2881 GTAAACGGCAAATTGAACGGTCAGGGAACATTCCGCTTTATGTCGGAACT 2930
                          ||||||   ::::::::::::
 823 .......................GlyThrThrGluIleValAsn..... 829

2931 CTTCGGCTACCGCAGCGACAAATTGAAGCTGGCGGAAAGTTCCGAAGGCA 2980
                    ||||||:::::::::|||::::::   :::
 830 ..........AlaHisAspLysValArgPheAlaAsnGlyLeuAsnThrL 843

2981 CTTACACCTTGGCGGTCAACAATACCGGCAACGAACCTGCAAGCCTCGAA 3030
          :::   |||:::   ::::::   :::      ::::::
 843 ysValSerAlaAlaThrValGluSerThrAspAlaAsnGlyAspLysVal 859

3031 CAATTGACGGTAGTGGAAGGAAAAGACAACAAACCGCTGTCCGAAAACCT 3080
           |||   |||:::          :::   ||||||::::::
 860 ThrThrThrPheValLysThrAspValGluLeuProLeuThrGlnIleTy 876

3081 TAATTTCACCCTGCAAAACGAACACGTCGATGCCGGCGCGTGGCGTTACC 3130
      |||                     :::|||||::::::   :::
 876 rAsn.....................ThrAspAlaAsnGlyLysLysIleT 886

3131 AACTCATCCGCAAAGACGGCGAGTTCCGCCTGCATAATCCGGTCAAAGAA 3180
          :::   |||||||||:::                    :::|||
 886 hrLysValValLysAspGlyGln..................ThrLysTrp 896

3181 CAAGAGCTTTCCGACAAACTCGGCAAGGCAGAAGCCAAAAAACAGGCGGA 3230
        ||||||:::          |||   |||:::      |||::::::
 897 TyrGluLeuAsnAlaAsp...GlyThrAlaAspMetThrLysGluValTh 912

3231 AAAAGACAACGCGCAAAGCCTTGACGCGCTGATTGCGGCCGGGCGCGATG 3280
           |||::::::|||                        |||:::   :
 912 rLeuGlyAsnValAspSer..................AspGlyLysLysV 923

3281 CCGTCGAAAAGACAGAAAGCGTTGCCGAACCGGCCCGGCAGGCAGGCGGG 3330
     ::|||:::   :::::::::            |||:::      |||
 923 alValLysAspAsnAspGlyLysTrpTyrHisAlaLysAlaAspGly... 938

3331 GAAAATGTCGGCATTATGCAGGCGGAGGAAGAGAAAAAACGGGTGCAGGC 3380
                        |||:::::::   |||   :::|||:::
 939 .................ThrAlaAspLysThrLysGlyGluValSerAs 949

3381 GGATAAAGACACCGCCTTGGCGAAACAGCGCGAAGCGGAAACCCGGCCGG 3430
      ||||||   ::::::      |||:::
 949 nAspLysValSerThrAspGluLysHis..................... 958

3431 CTACCACCGCCTTCCCCCGCGCCCGCCGCGCCCGCCGGGATTTGCCGCAA 3480
                                              :::   :::
 959 .....................................ValValSer 961

3481 CTGCAACCCCAACCGCAGCCCCAACCGCAGCGCGACCTGATCAGCCGTTA 3530
     |||:::|||:::   |||   :::   :::      :::|||::::::
 962 LeuAspProAsnAspGlnSerLysGlyLysGlyValValIleAspAsnVa 978

3531 TGCCAATAGCGGTTTGAGTGAATTTTCCGCCACGCTC.........AACA 3571
      ||||||:::   :::|||      |||:::           ::::
 978 lAlaAsnGlyAspIleSerAlaThrSerThrAspAlaIleAsnGlySerG 995

3572 GCGTTTTCGCCGTACAGGACGAATTAGACCGCGTATTTGCC......... 3612
     ::::::::||||||         :::   ::::::   :::
 995 lnLeuTyrAlaValAlaLysGlyValThrAsnLeuAlaGlyGlnValAsn 1011

3613 ......GAAGACCGCCGCAACGCCGTTTGGACAAGCGGCATCCGGGACAC 3656
           |||   :::   |||   ||| |||:::   |||::::::
```

```
1012 AsnLeuGluGlyLysValAsnLysVal.GlyLysArgAlaAspAlaGlyT 1028

3657 CAAACACTACCGTTCGCAAGATTTCCGCGCCTACCGCCAACAAACCGACC 3706
          :::::::::          :::   |||   |||:::   |||
1028 hrAlaSerAlaLeuAlaAlaSerGlnLeuProGlnAlaThrMetPro... 1043

3707 TGCGCCAAATCGGTATGCAGAAAAACCTCGGCAGCGGGCGCGTCGGCATC 3756
                           |||:::   :::|||   |||:::::||
1044 .................GlyLysSerMetValAlaIleAlaGlySerSe 1054

3757 CTGTTTTCGCACAACCGGACCGAAAACACCTTCGACGACGGCATCGGCAA 3806
     |            :::|||                    ::::::      :
1054 rTyrGlnGlyGlnAsnGly......LeuAlaIleGlyValSerArgIleS 1069

3807 CTCGGCACGGCTTGCCCACGGCGCCGTTTTCGGGCAATACGGCATCGACA 3856
     ::   :::|||              :::||||||:::|||   |||
1069 erAspAsnGlyLysValIleIleArgLeuSerGlyThrThrAsnSerGln 1085

3857 GGTTCTACATCGGCATCAGCGCGGGCGCGGGTTTT 3891
     |||:::|||  |||:::::||||||:::|||:::
1086 GlyLysThr..GlyValAlaAlaGlyValGlyTyr 1096
```